Computer Simulations of Lipid Bilayers and Proteins

Sonne, Jacob

Publication date:
2006

Document Version
Early version, also known as pre-print

Ph.D. Thesis
Computer simulations of lipid bilayers and proteins

Jacob Sonne

Technical University of Denmark
Department of Chemistry

Lyngby 2006

Type	Ph.D. thesis
Text and layout	Jacob Sonne
E-mail	mrsonne@gmail.com
Thesis title	Computer simulations of lipid bilayers and proteins
Thesis submitted	April 2006 Department of Chemistry Technical University of Denmark (DTU), Kgs. Lyngby, Denmark
Defense & approval	June 2006
Supervisors (DTU)	Günther H. Peters Flemming Y. Hansen Morten Ø. Jensen
Supervisor at the University of Calgary	D. Peter Tieleman
Examiners	Sievert-Jan Marrink John H. Ipsen Klaus B. Møller
This version	August 2007 Printed by Frydenberg A/S Copenhagen, Denmark
ISBN	87-91233-14-3
EAN	9788791233142

Preface & Acknowledgments

This thesis is submitted as partial fulfillment of the requirements for obtaining the Ph.D. degree at the Technical University of Denmark (DTU). The work was carried out from April 1st 2003 to March 31st 2006 mainly at the Department of Chemistry, DTU at MEMPHYS - Center for Biomembrane Physics under the supervision of Günther H. Peters, Flemming Y. Hansen and Morten Ø. Jensen. From February 2005 to August 2005 I visited the Department of Biological Sciences at the University of Calgary in the Biocomputing Group under the supervision of D. Peter Tieleman.

I would like to express my gratitude to my supervisors Günther Peters, Flemming Hansen and Morten Jensen for providing a professional, yet, informal environment nourishing scientific curiosity. From the MEMPHYS center I would like in particular to thank John Ipsen and fellow Ph.D. students Ask Jakobsen as well as Thomas Pedersen Nygaard for valuable discussions and of course thank the center director Ole Mouritsen for giving me the opportunity to study lipid bilayers at MEMPHYS - a Center of Excellence.

I am also truly thankful for Peter Tieleman granting me the opportunity to visit his wonderful group on the Canadian prairie, where I spent seven months in the interesting realm of ABC transporters, cowboys and grizzly bears. Thanks to all the people in the Tieleman lab as well as their special friends for being welcoming and helpful.

Last, but not least, I would like thank Olle Edholm and John Nagle for advice and suggestions.

Summary

The importance of computer simulations in lipid bilayer research has become more prominent for the last couple of decades and as computers get even faster, simulations will play an increasingly important part of understanding the processes that take place in and across cell membranes.

This thesis entitled *Computer simulations of lipid bilayers and proteins* describes two molecular dynamics (MD) simulation studies of pure lipid bilayers as well as a study of a transmembrane protein embedded in a lipid bilayer matrix. Below follows a brief overview of the thesis.

Chapter 1, Introduction: This chapter is a short introduction, where I briefly describe the basic biological background for the systems studied in Chapters 3, 4 and 5. This is done in a non-technical way to allow the general interested reader to get an impression of the work.

Chapter 2, Methods: In this chapter the background for the methods used in the succeeding chapters is presented. Details on system setups, simulation parameters and other technicalities can be found in the relevant chapters.

Chapter 3, DPPC lipid parameters: The quality of MD simulations is intimately dependent on the empirical potential energy function and its parameters, i.e., the force field. The commonly employed CHARMM27 force field reproduces fluid phase bilayer properties only when a tension is applied in the simulations. Without this tension the simulated bilayers become too ordered. To allow for more realistic lipid bilayer simulations in the tensionless isothermal-isobaric (NPT) ensemble, we reparameterized a part of the CHARMM27 force field by assigning new partial charges to the lipid head group atoms. Our modified CHARMM27 force field was tested in lipid bilayer MD simulations and was found to improve the phase properties of the simulated bilayers significantly. Thus, the improved force field makes it possible to simulate the biologically relevant fluid (L_α) phase in an NPT ensemble, which is an important prerequisite for taking full advantage of the predictive power of MD simulations since the area per lipid need not be known prior to simulation.

Chapter 4, Pressure profile calculations in lipid bilayers: A lipid bilayer is merely $\sim$5 nm thick, but the lateral pressure (parallel to the bilayer plane) varies several hundred bar on this short distance (normal to the bilayer). These variations in the lateral pressure are commonly referred to as the pressure profile. The pressure profile changes when small molecules partition into the bilayer and it has previously been suggested that such changes may be related to general anesthesia. MD simulations play an im-

portant role when studying the possible coupling between general anesthesia and changes in the pressure profile since the pressure profile cannot be measured in traditional experiments. Even so, pressure profile calculations from MD simulations are not trivial due to both fundamental and technical issues. We addressed two such issues namely the uniqueness of the pressure profile and the effect of neglecting pressure contributions from long range electrostatic interactions. The first issue is addressed by comparing two methods for calculating pressure profiles, and judged by the similar results obtained by these two methods the pressure profile appears to be well-defined for fluid phase lipid bilayers. For the second issue, we developed a method that allows pressure contributions from long range electrostatic interactions to be included in pressure profile calculations.

Chapter 5, Structural transitions in the ABC transporter BtuCD: In this chapter, we present a study of the transmembrane protein BtuCD. BtuCD belongs to the adonesine triphosphate (ATP) binding cassette (ABC) transporter family that use ATP to drive active transport of a wide variety of compounds across cell membranes. BtuCD accounts for vitamin B_{12} import into *Escherichia coli* and is one of the only ABC transporters for which a reliable crystal structure of the whole transporter has been determined. The (dys)function of ABC transporters accounts for cystic fibrosis and multi-drug resistance, for example tumor cell resistance to anticancer drugs. Hence, the mechanism facilitating substrate translocation in ABC transporters is of both fundamental and medical interest. It is commonly accepted that substrate translocation relies on binding and hydrolysis of ATP and maybe also on the release of hydrolysis products; however, the global structural rearrangements induced by these events remain largely unknown. To investigate these structural rearrangements in BtuCD we employed perturbed elastic network calculations and biased MD simulations. Comparing the results of these calculations with two transport models proposed in the literature, we are able to favor one over the other. Our observations for BtuCD may be relevant for all ABC transporters, owing to the conservation of ATP binding domains and the shared role of ATP in ABC transporters.

Chapter 6: This chapter contains a more technical summary of the thesis and some general conclusions.

Chapter 7: This chapter contains a compilation of supplementary material relating to the subjects discussed in the preceding chapters.

Contents

Introduction 1

1.1 Life, model systems & computer simulations

1.1.1 Cells & biological membranes

All living organisms consist of cells. Single cell organisms, which are arguably the smallest units of life, include bacteria and achaea, while multicellular organisms, such as plants and animals, consist of billions of cells. Fig. 1.1(b) shows an artist's rendering of a typical animal cell which could come from the grazing wapitis in panel (a). The shape, size and composition of cells vary according to their function, but one ubiquitous motif is the plasma membrane. Plasma membranes consist of different phospholipids arranged in a bilayer, which encapsulates the cytoplasm. In eukaryotic cells, like the one shown in Fig. 1.1(b), the cytoplasm contains different functional sub-compartments known as organelles. The bilayer matrix of the plasma membrane contains vital components such as cholesterol or hopanoids as well as integral proteins such as the one shown in Fig. 1.1(d). Additionally, the surface of some plasma membranes are decorated with glycolipids and glycoproteins [cf. Fig. 1.1(c)]. Thus, the plasma membrane is a quite complex system, a "mosaic" of different molecules, and owing to lateral fluidity of the bilayer matrix it is commonly referred to as a "fluid mosaic" [133].

1.1.2 Phospholipids & bilayers

Glycero phospholipid bilayers are commonly used as model systems for fully dressed biological membranes and have been studied intensively in experiments (see for example Refs. [115, 116, 124, 132]). Glycero phospholipids, or just phospholipids, consist of a triglycerol backbone where the first and second carbons each are attached to an aliphatic chain through an ester linkage. The third carbon is bound to a polar phosphate derivative which is referred to as the lipid head group. Fig. 1.2(a) and (b) illustrate lipids with the phosphatidylethanolamine (PE) and phosphatidylcholine (PC) headgroups, respectively. In bilayers, the aliphatic acyl chains of each monolayer face each other, while the head groups face the aqueous phase on both sides of the hydrophobic core. In Fig. 1.2(c) we have shown a bilayer

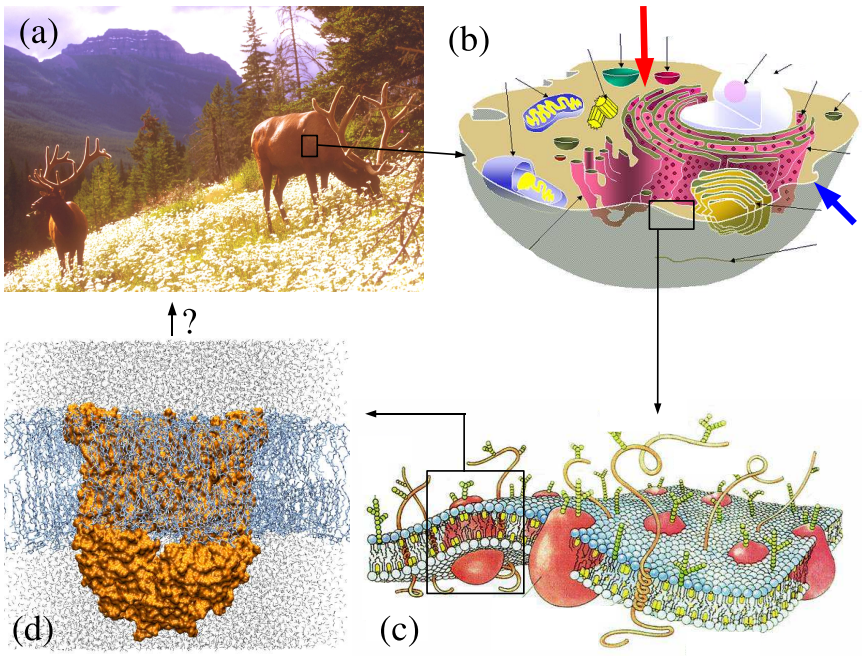
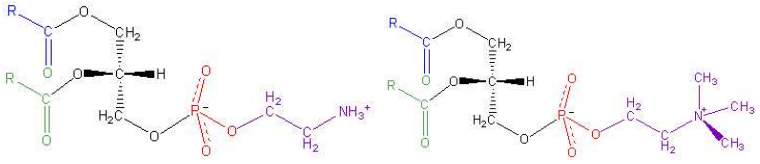


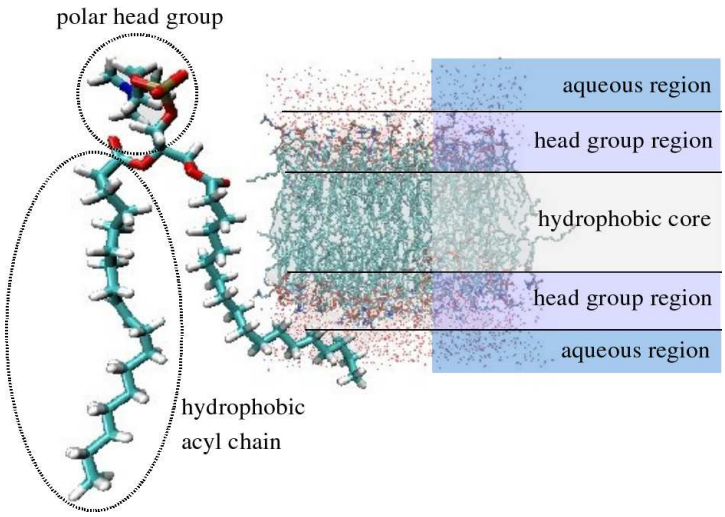
Figure 1.1: Different levels of viewing life. (a) Two wapitis grazing Leucanthemum near Banf, Alberta, Canada (photograph by Jacob Sonne) (b) Illustration of a typical animal cell. The red arrow points to the cytosol part of the cytoplasm while the blue arrow points at the plasma membrane. The thin black arrows point to different organelles (modified from image by Magnus Manske `http://en.wikipedia.org/wiki/Biological-cell`). (c) Plasma membrane (illustration by Dana Burns [23]). (d) Snapshot from a molecular dynamics (MD) simulation of an integral protein (orange surface representation) sitting in a lipid bilayer (blue licorice) surrounded by water (gray angles). The illustration was prepared using VMD [73].

patch and schematically illustrated the aqueous region(s), the head group regions, and the hydrophobic core region. To supplement the schematic lipids in Fig. 1.2(a) and (b), panel (c) shows a typical PC lipid configuration from a computer simulation of a lipid bilayer. The two acyl chains in a phospholipid need not be identical; their length and degree of saturation may vary. The lipid palmitoyloleonylphosphatidylethanolamine, for example, has one double bond in one of the acyl chains which is 18 carbon atoms long (oleonyl). The other chain, palmitoyl, is fully saturated and contains 16 carbon atoms. A compact notation for palmitoyloleonylphosphatidylethanolamine is (16:0/18:1c9)-PE or simply POPE. POPE is the lipid used in Chapter 5. Dipalmitoylphosphatidylcholine [(16:0/16:0)-PC or DPPC] is probably the most well-characterized lipid and is used in Chapters 3 and 4.

Above the main phase transition temperature T_m, lipid bilayers are in the fluid (L_α or liquid crystalline) phase, which is the physiologically relevant state of a pure bilayer [116, 146]. In the fluid phase trans-gauche isomerization occur relatively frequently ($\sim$0.1 ps^{-1}) and the acyl chains



(a) (b)



(c)

Figure 1.2: Lipid structures. (a) Schematic PE lipid moieties: glycerol (black), ethanolamine (violet), phosphate (red), The Rs (green and blue) symbolize the aliphatic acyl chains. (b) Schematic PC lipid moieties: Same as (a) except that ethanolamine is replaced by choline (violet). (c) Snapshot of a PC lipid from a MD simulation of a lipid bilayer. The whole bilayer is shown in the background with an indication of different regions.

are disordered. Lateral (in-plane) diffusion is fast, and two lipids would swap positions on a time scale of tens of nanoseconds [4, 121, 146] At temperatures below T_m, lipid bilayers are in the gel ($L_{\beta'}$) phase. Here, the acyl chain are highly ordered, tilted and lateral diffusion is slow. Below T_m, bilayers can also exist in a ripple phase and a crystalline phase, but these are not relevant for this thesis and will not be considered further.

1.1.3 Computer simulations of bilayer systems

The importance of computer simulations in lipid bilayer research has increased for the last couple of decades (see for example Refs. [115, 146, 150]). In molecular dynamics (MD) simulations of pure bilayers, DPPC is often

used and is considered a benchmark lipid (see for example [4, 51, 82, 98, 137, 143, 158]). POPE bilayers are commonly used to mimic the outer membrane of *Escherichia coli* (*E. coli*) [141, 142, 144], but should represent the plasma membrane equally well [118]. Most MD simulations have focused on the biologically relevant fluid (L_α) phase although a few MD simulation studies have (deliberately) simulated gel phase bilayers [92, 154]. Obtaining realistic fluid phase properties of pure lipid bilayers in MD simulations is far from trivial, but is a prerequisite for studying more complex bilayer systems. In Chapter 3 we address this problem.

For the past decades, the models and algorithms in MD simulations have improved and in turn made it possible to simulate large and complex systems. Now, bilayers doped with fatty acids and cholesterol [33], surfactants [12] and hexane [106] have been studied. Carrillo-Tripp and Feller [33] showed that when cholesterol molecules partition into the membrane, the distribution of lateral pressures (the pressure profile) in the membrane changes. Similar changes in the pressure profile have been suggested to be related to general anesthesia [30, 32]. Pressure profiles cannot be measured experimentally and therefore they have to be evaluated from computer models such as MD simulations. There are several fundamental and technical issues associated with pressure profile calculations from MD simulations, some of which are addressed in Chapter 4.

MD simulations of membrane proteins in atomistic detail are also accessible and such studies can provide unique insights into the details of protein function and protein-bilayer interactions [5, 79]. In Chapter 5 we focus on the integral protein BtuCD which is a so-called adenosine triphosphate (ATP) binding cassette (ABC) transporter. This family of transporters uses ATP for driving direct active transport of a wide variety of specific compounds and so there are 80 different ABC transporters in the gram-negative bacterium *E. coli* and 48 in humans [40, 71]. BtuCD is responsible for importing vitamin B_{12} into *E. coli* [103]. The relevance of ABC transporters in health research is dramatically illustrated by the fact that the ΔF508 mutation (phenylalanine deletion) in the cystic fibrosis transmembrane conductance regulator (CFTR) accounts for most cystic fibrosis cases [44, 126]. ABC transporters are also responsible for multidrug resistance [22], for example human tumor cell resistance to cytotoxic drugs used in chemotherapy [59, 72]. The motivation for studying an ABC transporter such as BtuCD using MD simulations, is that experiments have not yet fully uncovered the mechanism by which they facilitate transport. Understanding this mechanism has fundamental interest and could potentially be important in drug design.

Methods 2

2.1 Molecular dynamics simulations

In molecular dynamics (MD) simulations of bio-systems, the need for speed encourages most scientists to use highly optimized software packages such as AMBER [123], CHARMM[24, 108], GROMACS [18, 100] and NAMD [89], rather than writing the simulation code themselves. In the following, some of the background behind MD simulations is outlined. A more complete description of some of these subjects can be found in Refs. [64] and [2].

2.1.1 Time propagation

The total instantaneous total energy H in a constant energy simulation containing N particles reads

$$H = K + U(\mathbf{r}). \tag{2.1}$$

$\mathbf{r}$ is a $3N$ dimensional vector describing all atomic positions and $U(\mathbf{r})$ is the potential energy evaluated from the potential energy function (see Sec. 2.1.2). K is the kinetic energy given by $K = \sum_{i=1}^{N} m_i \mathbf{v}_i^2/2$, where $\mathbf{v}_i$ and m_i are the velocity and mass of the i'th particle, respectively. The instantaneous temperature T can be calculated from $K = 3Nk_BT/2$ if no constraints are used in the simulation. k_B is Boltzmann's constant. Providing initial positions and velocities for all particles i, allows for integrating Newton's equation of motion $\mathbf{f}_i = -\boldsymbol{\nabla}_{\mathbf{r}_i} U(\mathbf{r}) = m_i d^2(\mathbf{r}_i)/dt^2$ giving the time dependence of particle positions ($\mathbf{r}$), velocities ($\mathbf{v}$) and forces ($\mathbf{f}$).

There exists a host of integration algorithms designed for integrating the equations of motion and they have different merits and drawbacks. One choice is the Verlet half-step (leap-frog) algorithm, which propagates the position $\mathbf{r}_i$ and velocity $\mathbf{v}_i$ of atom i from time t to $t + \delta t$ according to the scheme:

$$\begin{aligned} \mathbf{v}_i(t + \delta t/2) &= \mathbf{v}_i(t - \delta t/2) + \mathbf{f}_i(t)\delta t/m_i \\ \text{and} & \\ \mathbf{r}_i(t + \delta t) &= \mathbf{r}_i(t) + \mathbf{v}_i(t + \delta t/2)\delta t. \end{aligned} \tag{2.2}$$

$\mathbf{f}_i$ is the force acting on atom i and is evaluated from the potential energy function. The velocity at time t is obtained as $(\mathbf{v}_i(t+\delta t/2)+\mathbf{v}_i(t-\delta t/2))/2$ and the total instantaneous energy of the system at time t can be calculated using Eq. (2.1). In, e.g., isothermal-isochoric or isothermal-isobaric simulations, the expression for the total energy and the integration scheme become more involved.

2.1.2 The potential energy function

The most central part of an MD simulation is arguably the potential energy function and its parameters. This combination is often referred to as the force field. The force field describes the multi dimensional potential energy surface of the system, and so for any configuration $\mathbf{r}$, the potential energy function gives the total configurational energy of the system. Note that electrons are not accounted for explicitely and therefore MD simulation is a classical approach.

As an example the CHARMM all-atom energy function [24] which was used for the bilayer studies in Chapters 3 and 4 is presented. The energy function consists of seven terms each accounting for different types of interactions:

$$\begin{aligned} U(\mathbf{r}) &= U_{\text{bond}} + U_{\text{angle}} + U_{\text{UB}} + U_{\text{dihedral}} + U_{\text{improper}} + U_{\text{LJ}} + U_{\text{C}} \\ &= \sum_{\text{bonds}} K_b(b-b_0)^2 \\ &+ \sum_{\text{UBs}} K_s(s-s_0)^2 \\ &+ \sum_{\text{angles}} K_\theta(\theta-\theta_0)^2 \\ &+ \sum_{\text{dihedrals}} K_\chi(1+\cos(n\chi-\delta)) \\ &+ \sum_{\text{impropers}} K_\phi(\phi-\phi_0)^2 \\ &+ \sum_{\text{non-bonded pairs } ij} \epsilon_{ij}\left(\left(\frac{r_{ij}^{\min}}{r_{ij}}\right)^{12} - 2\left(\frac{r_{ij}^{\min}}{r_{ij}}\right)^{6}\right) \\ &+ \sum_{\text{non-bonded pairs } ij} \frac{q_i q_j}{4\pi\epsilon_0\epsilon_r r_{ij}}. \end{aligned} \tag{2.3}$$

For clarity we have omitted full indexing in the summations, and the curious reader may find a more thorough description in Refs. [24, 109]. The bond energy U_{bond} is calculated as a sum over covalent bonds in the system. Each bond is treated as a harmonic spring with a specific force constant K_b. b is the actual bond length and b_0 is the equilibrium length of that specific

bond. Note that the harmonic potential does not allow bonds to break. U_{angle} is the valence angle energy, which increases harmonically when the angle between two neighboring covalent bonds θ departs from its equilibrium value θ_0. The Urey-Bradley term U_{UB} is a harmonic spring connecting atoms separated by two covalent bonds. s is the distance between those two atoms and s_0 is their equilibrium distance. Not all atoms separated by two covalent bonds are assigned a Urey-Bradley type interaction. A part of the torsional energy surface is accounted for by the dihedral interaction energy U_{dihedral}. When four atoms (a1-a2-a3-a4) are connected by three successive covalent bonds (b_{12},b_{23},b_{34}) without branches, χ is the angle between the two planes spanned by the b_{12} and b_{23} bond vectors and the b_{23} and b_{34} bond vectors. n is the symmetry of the rotor i.e. the number of minima for a 360° rotation in χ and δ is the phase. As a brief example, consider the torsional rotation around the central C2-C3 bond in butane, which has a global energy minimum in the *trans* configuration ($\chi = 180°$), whereas the *gauche* configurations ($\chi = \pm 60°$) have a higher energy. The highest energy is found in the *cis* configuration ($\chi = 0°$). The total dihedral energy for this rotation has contributions from three types of dihedral interactions, namely C-C-C-C, H-C-C-H and H-C-C-C. The contribution from each of these dihedral types may be a sum of multiple terms of the form $K_\chi(1 + \cos(n\chi - \delta))$. This sum of cosines sets the approximate locations of minima and maxima on the torsional potential surface, but does not appropriately lift the *gauche* energy relative to the *trans* energy. The increase in the *gauche* energy is introduced by non-bonded (U_{LJ} and U_{C}) interactions primarily between C1 and C4. U_{improper} is the energy of the improper, out-of-plane, torsions. The potential is harmonic and is for example used to assure a planar configuration around the ester linkages in phospholipids. U_{LJ} models van der Waals interactions between pairs of atoms through a 12-6 Lennard-Jones (LJ) potential. r_{ij} is the distance between atoms i and j and the parameter $r_{ij}^{\min}$ is the distance in which the LJ energy for pair ij attains its minimal value $-\epsilon_{ij}$. The LJ parameters for the pair ij are obtained by the Lorentz-Berthelodt combination rules: $r_{ij}^{\min} = (r_i^{\min} + r_j^{\min})/2$ and $\epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j}$. Electrostatics interactions between partial point charges q are evaluated from Coulomb's law. ϵ_0 is the permittivity of vacuum and ϵ_r is the relative dielectric constant. In all-atom MD simulations a value of 1 is usually adopted for ϵ_r, since all atoms are present explicitely and so electrostatic screening should be accounted for directly. The force field does not explicitely account for electronic polarization, and so ϵ_r should in principle be higher than 1 to emulate this effect. However, instead of adjusting ϵ_r, the polarization effect is incorporated into the partial charges [56, 107]. A similar approach has been used in other force fields [49, 143]. Usually non-bonded interactions (Lennard-Jones and Coulomb) are only evaluated between atoms separated by more than two covalent bonds.

2.1.3 Technicalities

Force fields

The CHARMM force field has been parameterized several times to reproduce experimental data for water, ions, alkanes, DNA, proteins and lipids [24, 51, 56, 94, 107, 130]. CHARMM22 and later versions represent all atoms explicitely and should always be used with the TIP3P water model [87, 107]. In Chapters 3 we describe a reparameterization of the CHARMM force field that aims at obtaining more realistic fluid phase bilayer properties in MD simulations at ambient pressures. In Chapter 5 we use the GMX3.2.1 ffgmx force field distributed with the GROMACS suite [18, 100]. The protein force field is based on a modified GROMOS87 force field [151]. Water is represented by the simple point charge (SPC) model [17]. Lipids are described by GROMOS bonded parameters, head group LJ parameters from OPLS [88] and special LJ parameters in the acyl chains [19]. Polar hydrogens in lipids and proteins are explicit, while non-polar groups containing hydrogen are represented by united atom particles. The united atom representation for the methyl(ene) groups in the lipid acyl chains significantly reduces the number of force evaluations and therefore gives a tremendous speedup compared to simulations with all-atom lipids.

Non-bonded interactions

The long range nature (r^{-1}) of Coulomb interactions indicates that these require extra care in MD simulations. Indeed several studies have documented structural artifacts in bilayers when using different cutoff-based strategies for evaluating electrostatic interactions [122, 154, 158].

In Ewald summation techniques, the electrostatic potential is not truncated and therefore such artifacts are not introduced, however, for large systems the method is computationally expensive. The Particle Mesh Ewald (PME) summation technique [39, 50] has almost linear scaling properties for large systems ($N \log N$) and has similar precision to the standard Ewald method if the grid spacing in the interpolation mesh is below 1 Å. Wohlert and Edholm [158] showed that the area per lipid is insensitive to system size and hydration level when using PME which is not the case when using an electrostatic cutoff. Thus, PME summation is the preferred method for evaluating electrostatic interactions in bilayer simulations. The Ewald summation is treated in some detail in Chapter 4, where we resolve a technical problem introduced when calculating pressure profiles from simulations conducted with Ewald summation.

Lennard-Jones interactions are treated with a cutoff in the range 1.0 to 1.4 nm. To avoid numerical problems associated with a truncated (discontinuous) potential, the Lennard-Jones 12-6 potential is often modified by a switch/shift function that assures that the energy smoothly goes to zero

when approaching the cutoff.

Macroscopic boundary conditions - ensembles

In the literature there has been an ongoing debate on, which ensemble is the most appropriate in lipid bilayer simulations. Setting up realistic MD bilayer simulations in the canonical ensemble NVT, where the number of particles N, the temperature T and the simulation box volume V is constant, is difficult since the correct box dimensions have to specified [121, 143]. In the isothermal-isobaric NPT ensemble, the simulation box adjusts its volume so that the internal pressure balances the external pressure P. In this thesis, the NPT-notation indicates that the simulation box length in the direction normal to the bilayer z is coupled to the barostat independently from the lateral directions x and y. Further $P_x = P_y = P_z$ and therefore the surface tension γ is zero [53]. As explained by Berger et al. [19], flaccid bilayers in experiments are able to adjust their area per lipid A. Minimizing the contact between the acyl chains and water and maximizing the acyl chain entropy creates a free energy minimum, which together dictate the equilibrium area per lipid A_{eq}, i.e., $(\partial G/\partial A)_{A=A_{\mathrm{eq}}} = \gamma = 0$. Therefore, it has been argued that the (tensionless) NPT ensemble is appropriate for simulating lipid bilayers [19, 77, 110].

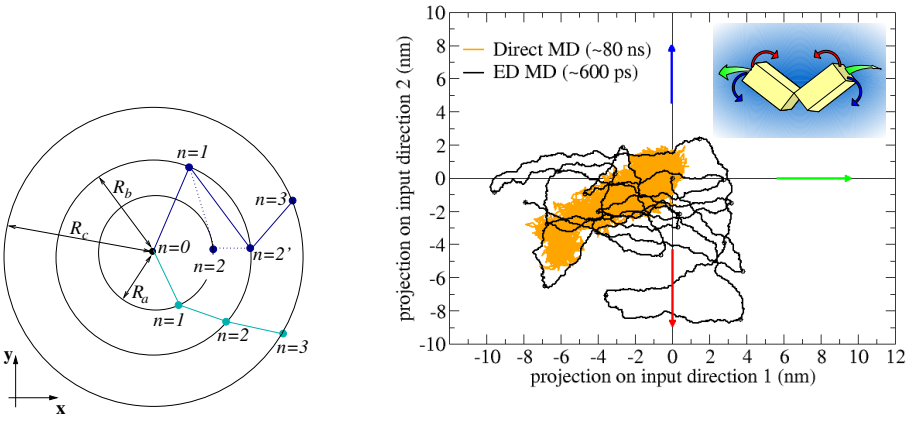
Bilayer simulations with a non-zero γ are either carried out in the $N\gamma P_z T$ ensemble where γ is specified directly or in the $NA_{xy}P_{\mathrm{z}}T$ ensemble where the simulation box xy-area A_{xy} is specified. These two ensembles have been used in bilayer simulations since it has been argued that a positive surface tension should be applied when simulating small bilayer patches [37, 51, 52, 53]. The applied surface tension in an MD simulation is given by $\gamma = \langle L_z(P_z - P_{\mathrm{L}})\rangle$ where P_z is the pressure in the direction normal to the bilayer, P_{L} is the average pressure in the lateral directions (x and y) and L_z is the z-dimension of the simulation box. Thus, applying a positive tension stretches the bilayer laterally and is thought for countering artificial lipid ordering in small bilayer patches where long wave length bilayer undulations are suppressed [53]. A theoretical study by Marsh [112] confirms this idea, but his estimate of the surface tension required to counter the artificial lipid ordering is smaller than the one suggested by Feller and Pastor [53]. More recently, Wohlert and Edholm [158] reported a relatively small finite size effect in MD simulations of a DPPC bilayer when using particle mesh Ewald summation for evaluating electrostatic interactions.

Hence, there seems to be little consensus in the ensemble discussion, but from a practical view, the NPT ensemble is more convenient since the area per lipid need not be known prior to simulation. This issue becomes more prominent when studying complex systems for example lipid mixtures, lipid bilayers with cholesterol and bilayers with embedded proteins.

2.1.4 Time scales and enhanced sampling

The time scales in lipid bilayers span 16 orders of magnitude from the bond vibrations ($\sim 10^{14}$ s^{-1}) to lipids flip-flopping between the two leaflets (~ 0.01 s^{-1}). To integrate the fast bond vibrations in MD simulations correctly, the equations of motion are usually propagated with a time step of one femtosecond. Therefore, accessing for example the flip-flop time scale is not possible. Constraining all bond lengths using e.g. LINCS [67] and SETTLE [113] eliminates the fastest modes of vibration, and therefore the time-step can be increased to ∼5 fs without affecting the equilibrium properties of the bilayer [4]. A speedup can also be obtained by using multiple time stepping schemes, where some interactions are updated less frequently than others [138]. At the present state of development, near-atomic detail lipid systems have been simulated for more than 100 ns, which is enough to observe self assembly of small vesicles from a random mixture of water and lipids [43] and to obtain reliable estimates for the lateral diffusion coefficient from pre-assembled bilayers [4, 99].

In protein simulations there is also a huge separation of time scales ranging from the fastest bond vibrations to the slow rearrangements in the tertiary structure. In between these regimes, changes in the secondary structure, such as helix unfolding and folding, has been simulated on a microsecond time scale in conventional (equilibrium) MD simulations [114]. For larger proteins, however, we cannot expect such an exhaustive exploration of the conformational space, partly because of the increased system size and partly because this would require sampling of slow structural deformations in the tertiary structure (global deformations). In Chapter 5 we investigate such global deformations in the ABC transporter BtuCD. To probe the large amplitude (∼long time scale) dynamics of the protein, devoid of sampling problems, we used MD simulations with essential dynamics sampling. In GROMACS [18, 100] different essential dynamics sampling algorithms are implemented. The input for all of them is a set of directions $\{\mathbf{x}\}$ in which the sampling should be increased. These directions are usually defined only for a subset of atoms for example all C_α atoms with initial positions $\mathbf{r}_0^{C_\alpha}$. Here we outline the fundamental idea behind two of these algorithms: 1. In each time step the atoms in the subset are displaced by a fixed increment in the input directions and the rest of the system is then relaxed in a regular MD integration step. In principle, this approach resembles a constant velocity steered MD simulation with an infinitely stiff spring attached to each atom in the subset. 2. After each regular MD step i, the displacement vector for the whole subset of atoms relative to the initial structure is projected onto the subspace span($\{\mathbf{x}\}$) and the norm of this projection R_i is calculated. If $R_i > R_{i-1}$ we have moved away from the initial structure and the MD step from $i-1$ to i is accepted. A new regular MD integration step brings the system to step $i+1$ where we check if $R_{i+1} > R_i$. If we, at some step k, find



(a) (b)

Figure 2.1: Enhanced sampling. (a) Schematic illustration of the expansive radial acceptance algorithm (see text for details). Two trajectories (cyan and blue lines) projected on the subspace spanned by the two input directions $\mathbf{x}$ and $\mathbf{y}$. The Rs are the distances between the structure at step n and the initial structure (n=0) in the span$(\mathbf{x}, \mathbf{y})$ subspace. (b) Two MD simulations of the periplasmic binding protein BtuF. The graph shows the trajectories projected onto two directions in the C_α subspace, which correspond to a opening/closing motion (blue/red) and a twisting motion of the two domains (green). The orange line is the projection of an 80 ns direct MD simulation (courtesy of Christian Kandt). The black line is the projection of a 600 ps simulation using expansive radial acceptance sampling (see text for details) where the expansion point was reset every 15 000th step. The inset schematically illustrates the opening, closing and twisting motions in BtuF.

that $R_k < R_{k-1}$, the positions of the atoms in the subset at step k are scaled uniformly so the new positions $(\mathbf{r}_{k'}^{C_\alpha})$ fulfill $R_{k'} = R_{k-1}$. The scaled coordinates are used for further propagation of the system. Algorithm 2 is known as acceptance based expansive essential dynamics sampling and is schematically illustrated in Fig. 2.1(a). The cyan trajectory starts at $n = 0$ and the regular MD integrator yields $R_{n=1} = R_a < R_{n=2} = R_b < R_{n=3} = R_c$ and so the essential dynamics algorithm does not interfere. The blue trajectory, on the other hand, contracts toward the initial structure from $n = 1$ to $n = 2$ $(R_{n=1} = R_b > R_{n=2} = R_a)$. Therefore, the expansive radial acceptance algorithm scales the positions at $n = 2$ to $n = 2'$ so that $R_{n=2'} = R_b$ is fulfilled. The $n = 2'$ positions are used for further propagation to $n = 3$. Protein configurations from trajectories, generated with these enhanced sampling algorithms, are not Boltzmann weighed, which makes the interpretation of averages and fluctuations non-trivial. Nevertheless, the algorithms are very appealing due to the effectiveness by which they sample the outskirts of the essential phase space. To illustrate this effectiveness, we simulated the aqueous protein BtuF in a water box. BtuF contains two domains in between

which there is a binding pocket specifically designed for binding vitamin B_{12} [91]. The opening/closing motion involving the two BtuF domains is functionally relevant for binding vitamin B_{12} in the periplasm of *E. coli* and releasing it into the ABC transporter BtuCD [101, 103]. Fig. 2.1(b) shows the dramatic increase in the sampling of the large amplitude motion associated with the functionally relevant opening/closing motion as well as the twisting motion in a 600 ps trajectory with expansive radial acceptance sampling compared to an 80 ns simulation with direct MD integration.

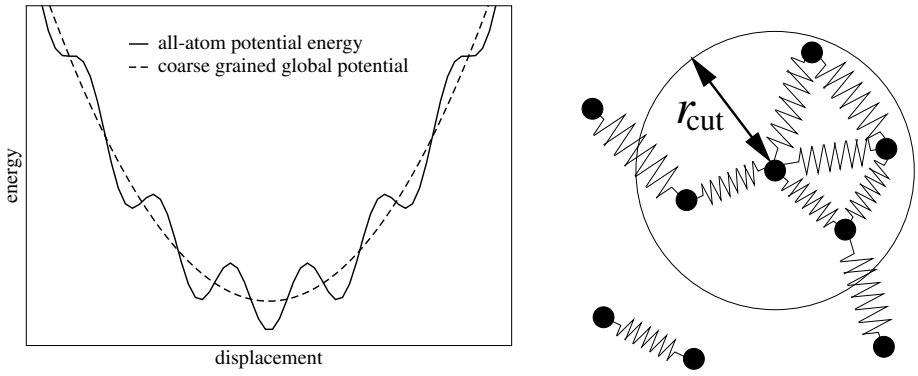
The input directions for the enhanced sampling algorithms can be taken from a set of experimentally determined protein structures, from a principal component analysis of a short direct MD trajectory or from coarse grained normal mode analysis [16, 84]. The essential dynamics sampling employed in Chapter 5 is based on experimentally determined protein structures.

2.2 Normal mode analysis

In a normal mode analysis (NMA) the multidimensional potential energy surface is expanded to second order around an energy minimum. The resulting matrix of second order derivatives, also known as the Hessian, is diagonalized which gives a set of eigenvalues and corresponding eigenvectors. Each eigenvector describes a direction in space along which the molecule (protein) vibrates with a frequency equal to the square root of the corresponding eigenvalue. Thus, modes with low eigenvalues correspond to motions taking place on long time-scales and in directions in which there is a low energy cost for deformation. Interestingly, it appears that proteins have evolved so that the directions of functionally important structural changes correspond to deformations in low frequency directions. In these directions, the intrinsic flexibility of the protein facilitates deformation at a low energy cost [105]. Thus, NMA is a useful tool for characterizing functionally relevant structural changes in proteins.

2.2.1 Elastic network models

In all-atom potentials, the large dimension of the Hessian matrix and the requirement that the protein structure should be in an energy minimum makes NMA computationally expensive for large proteins [16, 148]. Tirion [148] proposed that, in the folded state of a protein, detailed atom-atom interaction potentials can be replaced by harmonic potentials all with the same force constant. This approximation significantly reduces the computational burden and increases the stability of NMA for proteins. Tirion's one parameter model evolved into the so-called elastic network model [6, 7, 8, 9, 10, 70], where proteins are modeled as an elastic network of springs connecting C_α atoms within some cut-off [cf. Fig. 2.2(b)]. All atoms except C_αs are discarded from the calculation. This coarse graining smoothens the energy



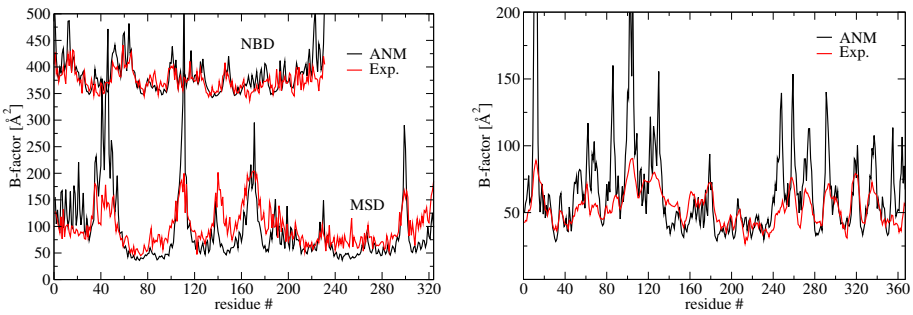
(a) (b)

Figure 2.2: Illustration of two aspects of the elastic network model. (a) Schematic illustration of an (all-atom) potential with local maxima and minima and the coarse grained (elastic network) global potential. (b) In the elastic network model C_αs within some cutoff distance r_{cut} are connected with a spring.

surface as illustrated in Fig. 2.2(a). Since local anharmonicities, that create the local maxima and minima in the all-atom potential, are not necessarily important for the global vibrations of the protein, the coarse grained potential may suffice for describing long time scale collective vibrations in a protein. In the elastic network model, the C_α-C_α springs are assumed to have their equilibrium length in the X-ray structure and are all assigned the same force constant γ. The same value for γ is used for all proteins and this universal value has been determined by fitting the magnitude of positional fluctuations of the C_αs in the network to experimental data (crystallographic B-factors). For a protein with N residues, diagonalization of the contact matrix H in the Anisotropic Network Model (ANM) yields $3N-6$ non-zero eigenvalues (Ω_i^2) with $3N-6$ corresponding eigenvectors ($\mathbf{u}_i$). The diagonalization allows us to write the potential energy of the network V^{en} in two equivalent forms:

$$V^{\text{en}} = \frac{\gamma}{2}\Delta\mathbf{R}_c\mathsf{H}\Delta\mathbf{R}_c^T = \frac{\gamma}{2}\Delta\mathbf{R}_u\Omega\Delta\mathbf{R}_u^T, \tag{2.4}$$

where $\Delta\mathbf{R}_c$ and $\Delta\mathbf{R}_u$ give the displacement from the crystal structure in Cartesian basis and eigenvector basis, respectively. Ω is a diagonal matrix with diagonal elements Ω_i^2 and from these, the vibrational frequency ω_i of mode i can be calculated as $\omega_i = \sqrt{\gamma}\Omega_i$. The eigenmodes from the ANM analysis can be visualized and correlations between residue displacements can be represented in two dimensional correlation maps. The elastic network model has repeatedly proved successful in reproducing crystallographic B-factors, which are readily available from the method [6]. As an example



(a) (b)

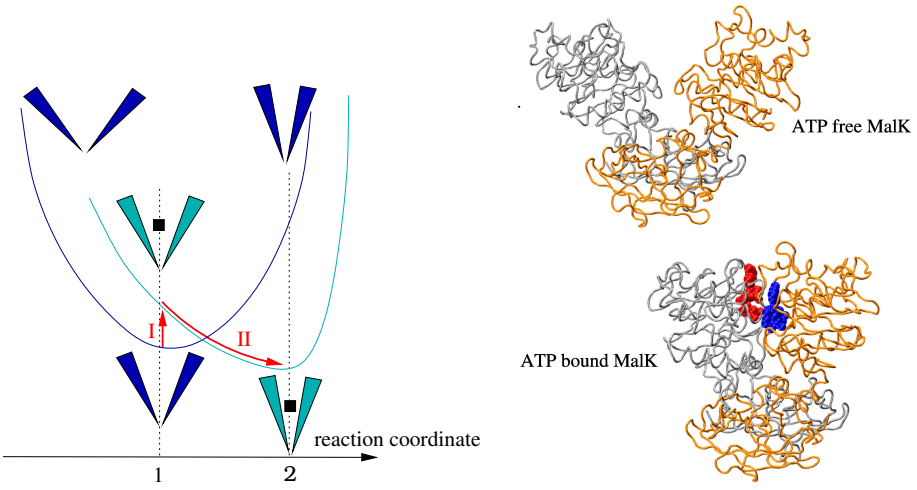
Figure 2.3: Comparison of crystallographic B-factors as determined in experiments (red) and in the anisotropic elastic network model (black) for (a) the membrane spanning domain (MSD) and nucleotide binding domain (NBD) parts of the ABC transporter BtuCD [103]. The B-factors of the NBDs are shifted by +300 Å^2 to increase clarity. The BtuCD structure is shown in Fig. 5.1 (Chapter 5.1); (b) MalK which is the NBD part of the maltose ABC transporter [36]. The MalK structure is shown in Fig. 2.4(b).

of the level of correspondence with experimental data, Fig. 2.3(a) shows the B-factors for the ABC transporter BtuCD and Fig. 2.3(b) shows the same for MalK, which is the nucleotide binding domain (NBD) part of the maltose ABC transporter. Overall, there is a good agreement between the theoretical and experimental B-factors for both proteins. Elastic network based modeling of these two proteins is an important part of Chapter 5. Note that changing the value of γ does not change the direction of the eigenvectors, only the amplitude of the vibrations. Consequently, the B-factors for all C_αs are scaled uniformly (B-factor $\propto \gamma^{-1}$). In addition to reproducing experimental B-factors, also estimates of H/D-exchange free energies and NMR order parameters have successfully been reproduced [11, 63]. Further, the low frequency modes predicted by the network have also been shown to be in good agreement with principal component analyses from MD simulations as well as normal mode analysis in all-atom force fields [47, 93]. On a standard PC the ANM calculation takes 10-20 hours for a $\sim 10^3$ residue protein such as the vitamin B_{12} transporter BtuCD.

2.2.2 Ligand binding & normal mode analysis

In elastic network calculations, and NMA in general, only vibrations around an energy minimum are predicted and non-equilibrium events such as ligand-induced structural deformations cannot be accounted for directly. Here, the term "ligand" covers any molecule that interacts with the protein triggering a structural transition. Since the intrinsic protein flexibility, provided by a few low frequent normal modes, often constitute the minimum energy

path between two functionally relevant states, NMA is nevertheless a useful tool for characterizing the directions that are functionally relevant [105]. This is illustrated in Fig. 2.4(a) where the schematic protein (dark blue) vibrates around the global energy minimum (1) in the ligand free state. In principle, it can access the (unstable) closed configuration (2) via vibrations in low frequent (equilibrium) normal modes. Ligand binding changes the energy landscape and so the global energy minimum shifts along the reaction coordinate so that the closed configuration (2) is stabilized. Fig. 2.4(b) shows an example of a ligand-induced structural transition in a real protein; in this case it is the ATP induced transition in MalK which plays a central role in Chapter 5.



(a) (b)

Figure 2.4: The effect of ligand binding. (a) Schematic illustration of changes in the free energy surface (along the reaction coordinate) and protein structure upon ligand binding. The dark blue protein structures and curve correspond to the ligand-free state while the cyan protein structures and curve correspond to the ligand-bound state. The ligand is symbolized by the filled black square. In the unbound state (dark blue) the average (minimum energy) configuration is denoted 1. The reaction coordinate is a linear combination of a few low frequent eigenmodes derived from NMA in the unbound minimum energy configuration (1). The red arrows indicate that when a ligand is introduced in configuration 1, (I) the energy surface changes from the dark blue to the cyan. On the cyan energy surface configuration 1 is no longer the minimum energy configuration and consequently (II) the protein relaxes to configuration 2, which is the new energy minimum structure. (b) ATP induced structural transition in MalK [36]. The two domains of MalK are rendered in orange and gray tubes, respectively. The two ATPs, which are bound in a sandwich between the two MalK domains, are rendered in blue and red van der Waals representation, respectively.

For determining the functionally relevant linear combination of low frequent eigenmodes from NMA, Zheng and Brooks [160] proposed a pertur-

bational variant of the elastic network. For the discussion of this method, let us denote the ligand-bound minimum energy configuration as the target state and the unbound minimum energy configuration as the initial state. In the perturbed elastic network approach it is assumed that the effect of ligand binding can be accounted for by knowing the distance changes of a few (say M) C_α pairs going from the initial to the target state. The target C_α pair distances can for example be obtained from experiments. The distance changes for the M C_α pairs going from the initial to the target state are introduced in the elastic network energy as M C_α-C_α harmonic distance constraints:

$$V = V^{\text{en}} + V^{\text{pert}} = V^{\text{en}} + \sum_{m=1}^{M} k_m^{\text{eff}} \left(d_{(ij)_m} - d_{(ij)_m}^{\text{target}} \right)^2 . \qquad (2.5)$$

$d_{(ij)_m}^{\text{target}}$ is the distance between C_αs i and j in the m'th residue pair in the target state and $d_{(ij)_m}$ is the current distance for that pair. k_m^{eff} is the effective force constant for the mth distance constraint and is given by $1/k_m^{\text{eff}} = \sum_{k=7}^{3N} \delta d_{(ij)_m}^k / \Omega_k^2$ where $\delta d_{(ij)_m}^k$ is the change in the pair distance $d_{(ij)_m}$ associated with mode k. The summation starts at $k = 7$, to exclude the six modes with zero eigenvalues. The M distance constraints perturb the elastic network energy, which in effect moves the energy minimum away from the crystal structure. The method then predicts the response to this perturbation, i.e., the direction of the relaxation of the whole protein structure in this modified energy landscape. This is in essence very similar to the illustration of the effect of ligand binding in Fig. 2.4(a). The response vector from the perturbed network approach is a linear combination of the unperturbed ANM eigenmodes, favoring low frequency modes, and therefore the response will have the character of a global rearrangement of the protein rather than a local deformation. Since the output is only one ($3N$ dimensional) response vector that describes how all C_αs in the protein respond to changing M C_α-C_α pair distances, the method resolves the problem of finding functionally relevant vibrations among $3N - 6$ eigenmodes.

Zheng and Brooks [160] tested the perturbed elastic network method on 22 proteins for which two crystal structures, that represent two conformational states, are known. They chose ten C_α pairs for the perturbation for which the C_α pair distance in initial and the target state lie on mutually opposite sides of the network cutoff. Among these pairs they, essentially, choose the ten pairs with the numerically largest distance change between the two structures. For most of the 22 proteins tested, the perturbed network approach predicts the experimentally determined structural transition better than any unperturbed eigenmode. Therefore, the method seems very promising for predicting the reactions coordinate between two functionally relevant configurations of a protein. In Chapter 5 we use the method for modeling the ATP induced structural rearrangements in MalK and BtuCD.

3 DPPC lipid parameters

Based on: J. Sonne, M.Ø. Jensen, F.Y. Hansen, L. Hemmingsen and G.H. Peters. A parameter study of dipalmitoylphosphatidylcholine lipids. Submitted.

3.1 Introduction

In molecular dynamics (MD) simulations of biological membranes, lipid bilayers are the most common model system, and with the increase in computer power, simulations of larger and more complex bilayer systems are becoming accessible [5, 12, 33, 106]. In order to represent bilayers accurately in MD simulations, it is important to have a well-parameterized force field in combination with appropriate macroscopic boundary conditions (ensembles). Although, these topics may seem somewhat technical in a biological context, the functional importance of cell membranes should not be overlooked. For example it has been shown that phospholipase activity change the mechanical properties of bilayers [26], which could have therapeutic application in delivery of cytotoxic anti-cancer drugs [3]. Further, bilayer properties affect mechanosensitive gating in McsL [38, 62, 139], phospholipase C activity [78] and have been suggested to be related to anesthesia [29, 32]. See Jensen and Mouritsen [79] for a recent review of bilayer influence on protein function.

Commonly used united atom force fields produce fluid (L_α) phase dipalmitoylphosphatidylcholine (DPPC) bilayers in the tensionless ($\gamma = 0$) isothermal-isobaric NPT ensemble provided that the correct electrostatic cutoff strategy is used [4, 122, 158]. In contrast, DPPC bilayers simulated using the all-atom CHARMM22 force field [107, 130] in the NPT ensemble show a dramatic lateral contraction and overly ordered lipid acyl chains [53]. Despite optimizations in both the head group and the acyl chains in the subsequent CHARMM27 parameter set [51, 56], the area per DPPC lipid molecule is underestimated by at least 0.15 nm^2 [82] compared to experimental data [117]. In fact, the area per DPPC lipid in NPT simulations [82] is close to that of the gel phase [116]. Recently, the CHARMM27 force field was further optimized [94], but the revised parameters for the acyl chains do not produce fluid-like bilayers in the NPT ensemble [135]. Thus, to mimic the

biologically relevant fluid phase, most bilayer simulations using CHARMM parameters apply a positive surface tension [33, 51, 53, 62, 82, 94, 95]. It has been suggested that the gel-like properties in NPT simulations can be attributed to finite size effects and therefore applying a positive surface tension, that stretches the bilayer to the experimentally determined area per lipid, is appropriate [51, 52, 53]. However, finite size effects seemingly account for an area contraction of less than 1 $Å^2$/lipid [158], yet, the area per lipid is underestimated by at least 15 $Å^2$ compared to experimental data for a DPPC bilayer (72 lipids) simulated in the NPT ensemble with CHARMM27 parameters [82]. Thus, even though MD simulations of bilayers are subject to finite size effects, these can by no means account for the gel-like bilayer properties in NPT simulations using CHARMM27 parameters. It appears that the gel-like properties are mostly a result of the force field not being optimized for lipid bilayer simulations in the NPT ensemble, but for simulations with an applied surface tension.

Although, there is no agreement on whether or not applying a surface tension is appropriate when simulating lipid bilayers [19, 51, 52, 53, 77, 110], tensionless NPT simulations are appealing from a practical view since the area per lipid need not be known prior to simulation, if the parameters are optimized for this ensemble. The advantage of having parameters that are optimized for NPT simulations becomes more prominent when studying for example lipid mixtures and bilayers with embedded proteins where experimental data is scarce. Therefore, realistic NPT bilayer simulations have been one ultimate goal for MD force fields [15, 94].

Along these lines, a reparameterization of the CHARMM phosphatidylcholine (PC) lipid parameters that permits simulations of fluid phase DPPC lipid bilayers in the NPT ensemble is the purpose of this study. The lipid head group charges is the target for our reparameterization. This chapter is structured as follows: In Sec. 3.2 we describe the parameterization strategy and in Sec. 3.3 we give the technical details of our calculations and simulations. We present and discuss our results in Secs. 3.4 and 3.5, respectively, and summarize our findings in Sec. 3.6. The partial charges resulting from our study are provided in Appendix 7.1.

3.2 Methods

We start this section by outlining relevant aspects of the CHARMM optimization strategy before we describe the methods used in this study.

3.2.1 CHARMM27 parameterization strategy

In the CHARMM force field electrostatic interactions are accounted for by partial point charges located at the atomic centers [24, 51, 56, 108, 130]. Since electrons are delocalized, this representation cannot be exact and there

is no rigorous way of determining such atomic charges. Nevertheless, finding suitable partial charges that, with reasonable precision, account for molecular properties is one task in parameterizations of this force field. Complying with the CHARMM strategy, Foloppe and MacKerell, Jr. [56] optimized the CHARMM partial charges of dimethylphosphate to reproduce quantum mechanical interaction energies with a strategically placed water molecule using a Mulliken population analysis to propose the initial partial charges. The optimized charges are now used for the phosphate moiety in the head group of phospholipids [51].

In the optimization of the CHARMM27 parameters, Foloppe and MacKerell, Jr. [56] distinguish between macromolecular target data and small molecule target data. The ability of the CHARMM27 force field to reproduce the former has the highest priority. In the case of lipid bilayers macromolecular target data would be bilayer properties such as lipid densities, electron densities and order parameters, whereas small molecule target data would be, e.g., torsional energy surfaces in the lipid headgroup and water-lipid interaction energies.

3.2.2 Our parameterization strategy

Pressure profile calculations for lipid bilayers suggest that the equilibrium area per lipid is determined by a delicate balance between large and opposing forces originating from bonded and non-bonded interactions [98, 137]. Therefore, reparameterization of any of the terms in the energy function could, in principle, affect the area per lipid. NPT and $NP_z\gamma T$ simulations of a crystalline (all-trans) C_{36} alkane showed good agreement with experiments [81, 83], which indicates that the alkane-alkane interactions are well parameterized. With the recent refinement of the torsional potential for CHARMM27 alkanes, the parameters for the acyl chains of phospholipids also seem highly optimized [94]. Our attention therefore turns to the lipid head group region. The inter-headgroup interactions are determined by the Lennard-Jones parameters and the partial charges. We believe the latter to be a more promising optimization target, since numerous united atom force field bilayer simulations have shown that the bilayer properties are quite sensitive to details in the treatment of the electrostatic head group interactions [4, 122, 158]. Moreover, pressure profiles derived from atomistic simulations, indicate that electrostatic attractions significantly contribute to the positive surface tension in the head group region. This indicates that the electrostatic forces in this region are, on average, contractive [98, 137]. Therefore, the lipid head group charges will be the target of our reparameterization.

We determine initial partial charges of the whole lipid head group and upper acyl chain from *ab initio* data using (I) a Mulliken population analysis and (II) a restricted electrostatic potential (RESP) fitting approach [14],

i.e., we adjust the partial atomic charges to fit the quantum mechanical electrostatic potential (see Sec. 3.3 for further details). Since the RESP charges generate a realistic electrostatic potential around the molecule of interest, the method is optimal for reproducing intermolecular interactions [14]. The RESP method has been shown to be superior to Mulliken population analysis [134, 159], but since determination of the Mulliken charges does not add any significant overhead to the calculation of the quantum mechanical electrostatic potential, we include both methods in the present study.

In the RESP method, the derived charges are known to be strongly dependent on the conformation of the molecule. Reynolds et al. [125] addressed this problem by deriving RESP charges from different conformers and then estimated the final set of partial charges as the Boltzmann weighted average of the charge sets found for the different conformers. We adopt a similar approach by extracting 69 dipalmitoylphosphatidylcholine (DPPC) configurations from a 15 ns MD simulation of a DPPC lipid bilayer, where the area was fixed at the experimental value of 62.9 $Å^2$ per lipid [117]. This experimental estimate was later adjusted to 64 $Å^2$ [116], but for extracting lipid conformations, this adjustment should be of minor importance. These 69 DPPC molecules were capped to form dipentanoatephosphatidylcholine, DPePC, shown schematically in Fig. 3.2. Since numerous studies have shown that the CHARMM alkane parameters are highly optimized [51, 81, 83, 94] we did not include the complete acyl chains in our *ab intio* calculations. From the 69 DPePC configurations we determined the Mulliken and RESP charges and calculated the final charges as a simple average over the conformers, which are already Boltzmann weighted from the MD simulation. Thus, our approach bypass the issue of the Boltzmann weights being different in vacuum and in condensed phases [14].

When changing the atomic partial charges, the remainder of the force field parameters could be readjusted iteratively to maximize accordance with, e.g., *ab initio* potential energy surfaces and vibrational data, i.e., small molecule target data [56]. However, our main motivation for assigning new partial charges to DPPC lipids is to obtain fluid-like properties of DPPC bilayers, i.e., to optimize the macromolecular (bilayer) properties only. Therefore, testing whether our new charges reproduce small molecule target data is beyond the scope of this work. Our new DPePC partial charges were therefore transferred to DPPC without modifying the force field further. We used two schemes for this transfer: (A) All charges from DPePC were used in DPPC except for the terminal C26 methyl groups (see Fig. 3.2). (B) All charges from DPePC were used in DPPC except for the terminal C24-C26 ethyl groups. Excluding the methyl and ethyl groups in schemes (A) and (B), respectively, leaves DPePC, and therefore also DPPC, with a non-zero net charge. To obtain a neutral DPPC molecule we compensated this net charge with small counter charges on each non-acyl atom in DPPC. We choose not to cluster our charges in charge groups, with in-

tegral values for the group charges, since most membrane simulations using the CHARMM force field evaluate electrostatic interactions by PME summation rendering charge clustering superfluous. The lack of charge groups abrogates transferability of our charges to other phospholipids such as phosphatidylethanolamines.

3.2.3 Testing the new parameters

Combining the Mulliken population analysis (I) and RESP (II) with transfer schemes (A) and (B) yields four sets of partial charges, which we subsequently tested in four MD simulations of a DPPC lipid bilayer. The simulation results, referred to as I.A, I.B, II.A and II.B, were compared with NPT and $NP_z\gamma T$ experimental data and with simulations using the CHARMM27 parameters.

As benchmark properties for the comparison we resorted to the volume and area per lipid, the order parameter profile and the electron density profile. The volume per lipid is calculated by subtracting the water volume $n_\mathrm{w}V_\mathrm{w}$ from the box volume [54], where n_w is the number of water molecules and V_w is the average volume of one bulk water molecule in the simulation. The area per lipid is obtained as the area of the simulation box xy-plane divided by the number of lipids in one leaflet. The order parameter for the ith acyl methyl(ene) group $|S_{\mathrm{CD},i}|$ is calculated as $|S_{\mathrm{CD},i}| = |\langle 3/2\cos^2\theta_i - 1/2\rangle|$ where θ_i is the angle between a C-H bond vector in the ith methyl(ene) group and the bilayer normal, i.e., the z-axis. The brackets denote averaging over the C-H bonds in the ith methyl(ene) group, lipids and time. The electron density profile was calculated by binning the difference $Z - q$ for all atoms along the z-axis, where Z is the atom number and q is the atomic partial charge. Thus, we neglect bilayer undulations and assume that the electrons are located at the atomic centers, thereby ignoring variations in the atomic form factors. Both assumptions are reasonable when simulating a small fluid bilayer patch [95].

3.3 Simulation details

3.3.1 Atomic partial charges

From an MD simulation of a DPPC lipid bilayer where the area was fixed at 62.9 Å^2 per lipid, we extracted 69 DPPC configurations after ∼15 ns (see Sec. 3.3.2). These 69 DPPC molecules were capped to form dipentanoate-phosphatidylcholine (DPePC) by (a) cutting off the lower decanyl parts of the two acyl chains, (b) replacing the last carbon (C_6) in the hexanoate acyl chains with hydrogen and (c) adjusting the new C_5-H bond length accordingly. Without geometry optimization the quantum mechanical electrostatic potential around each DPePC configuration was evaluated at the RHF/6-

31G(d) level in ∼100 000 points using Gaussian98 [57]. In addition to the electrostatic potential, we also calculated the atomic charges from a Mulliken population analysis. From the quantum mechanical electrostatic potential, atomic partial charges were determined by two successive restricted electrostatic potential (RESP) fits as described by Bayly et al. [14]. In the first fit we used the CHARMM27 charges as the initial guess and used no symmetry constraints. The charges were restrained by a hyperbolic penalty function with weight 0.0005 au to avoid large charge separation. In the successive fit, the output charges from the first fit were used as input, and the hyperbolic penalty weight was increased to 0.001 au. In the second fit we enforced symmetry constraints for equivalent atoms (see Fig. 3.2). The same definition of equivalent atoms was used to symmetrize the charges from the Mulliken population analyses. Lastly, we averaged partial charges over the 69 DPePC configurations to give the average partial charge distribution in the DPePC molecule from the Mulliken and RESP procedures, respectively. The RESP fitting was performed using Amber 4.1 with the Restrained ESP Fit package 2.3 [123].

3.3.2 MD simulations

The DPPC lipid bilayer, from which we extracted the lipid conformations for the atomic charge determination, consists of 72 lipids solvated with ∼2 000 water molecules resulting in a total of ∼16 000 atoms. This corresponds to ∼29 water molecules per lipid [117]. The water molecules were placed around the bilayer using Solvate [60] and subsequently the water layer surrounding the bilayer was cropped to a rectangular, periodic simulation box. The water molecules were represented by the TIP3 water model [87]. The bilayer system was equilibrated in the CHARMM27 force field for 15 ns with the area fixed at the experimental value, 62.9 Å^2 [117]. After this equilibration, we changed the partial charges to sets I.A, I.B, II.A and II.B in four simulations, which were then continued. We also carried out two reference simulations, referred to as III and IV, where the original CHARMM charges were used. For simulations I.A, I.B, II.A and II.B and III we used the NPT ensemble, thus allowing the three dimensions of the simulation box to adjust independently to the relevant target pressure components. Simulation IV was carried out in the $NP_z\gamma T$ ensemble with $\gamma = 61$ mN/m and $P_z = 1$ atm. In all simulations we used a time step of 1.0 fs and the target temperature of the Langevin thermostat was 325 K with a damping coefficient of 5 ps^{-1}. The pressure P =1 atm was controlled by the Nose-Hoover Langevin barostat [55] with a piston oscillation time of 100 fs and a damping time of 50 fs. Electrostatic interactions were evaluated using the PME method [39, 50] with a grid spacing below 1 Å. MD simulations were carried out using NAMD [89]

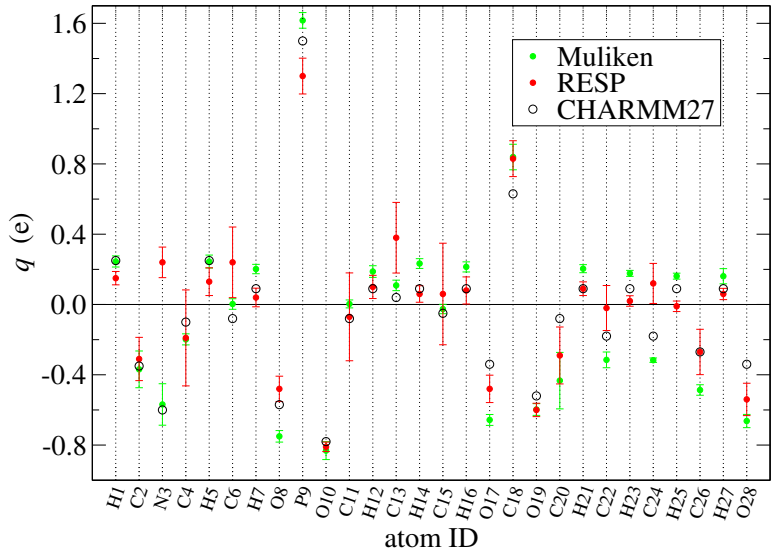
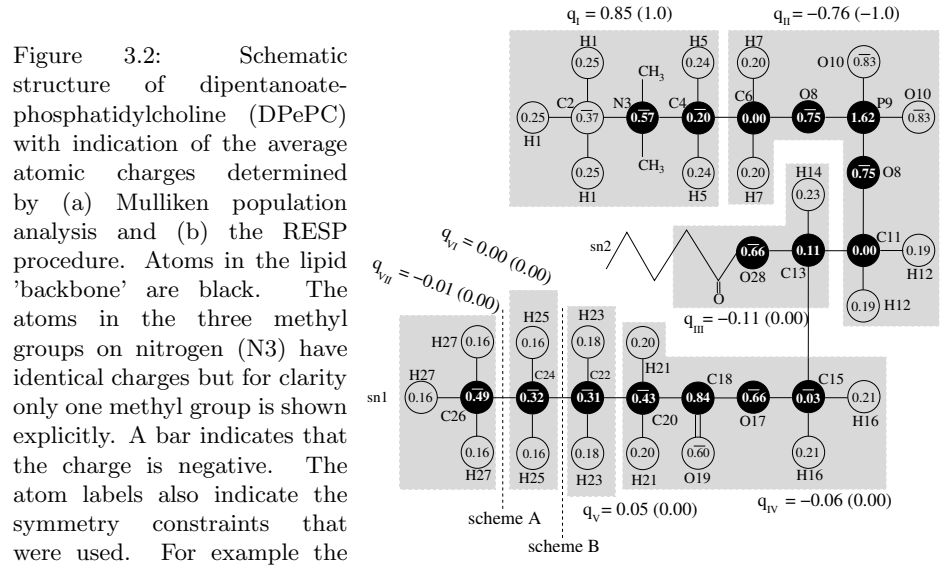


Figure 3.1: Overview of the obtained partial charges for dipentanoatephosphatidylcholine (DPePC). Average values and standard deviations for the new RESP charges (red) and new Mulliken charges (green). For comparison, we have included the charges from the CHARMM27 force field (black). The atom IDs used on the abscissa axis are defined in Fig. 3.2

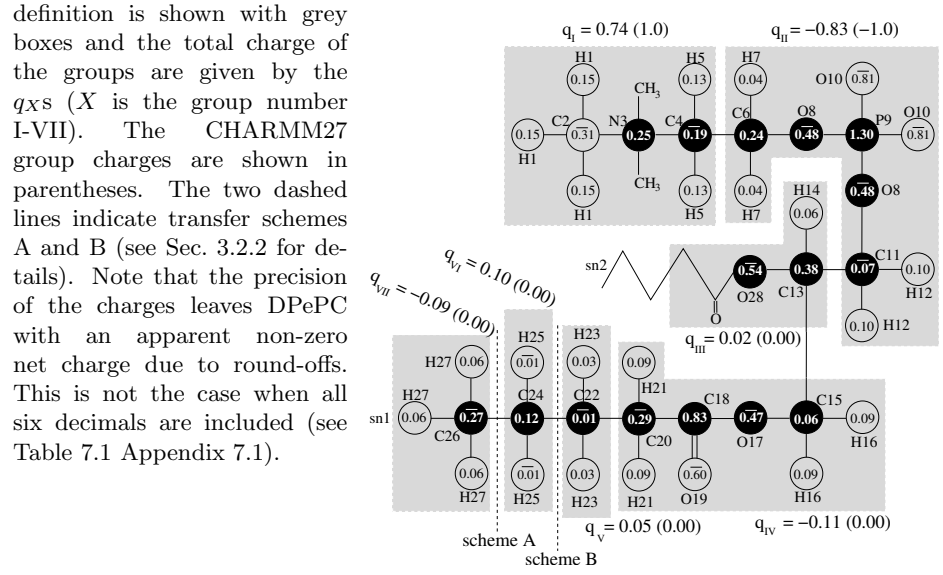
3.4 Results

3.4.1 Atomic partial charges

From the different configurations of DPePC we calculated the Mulliken and RESP charges for each configuration. In Fig. 3.1 we present average charges and standard deviations for Mulliken and RESP procedures. The charges are compiled in Table 7.1 in Appendix 7.1. To get a better overview of the charge distribution, the obtained charges are mapped onto DPePC in Fig. 3.2 with only two decimals, which leaves DPePC with a non-zero net charge due to round-off errors. With six decimals the new charges fulfill electroneutrality (cf. Table 7.1 in Appendix 7.1). Qualitatively, there is a good agreement between the CHARMM27 atomic charges and the charges obtained from the Mulliken method. This agreement is expected since the initial CHARMM partial charges were taken from a Mulliken population analysis [107]. In general the same is true for the obtained RESP charges, however, the RESP charges for N3, C6 and C24, have the opposite sign compared to the corresponding CHARMM27 charges. A similar sign inversion relative to CHARMM27 is not observed for our Mulliken charges and therefore this prediction appears to be specific for the RESP method. Comparing the error bars on the charges from the two methods, we see that the RESP charges exhibit a larger configurational dependence than the Mulliken charges. This is a commonly known difference between the two methods [14]



(a)



(b)

Figure 3.2: Schematic structure of dipentanoate-phosphatidylcholine (DPePC) with indication of the average atomic charges determined by (a) Mulliken population analysis and (b) the RESP procedure. Atoms in the lipid 'backbone' are black. The atoms in the three methyl groups on nitrogen (N3) have identical charges but for clarity only one methyl group is shown explicitly. A bar indicates that the charge is negative. The atom labels also indicate the symmetry constraints that were used. For example the two atoms labeled H5 have the same charge by definition. The CHARMM27 charge group definition is shown with grey boxes and the total charge of the groups are given by the q_Xs (X is the group number I-VII). The CHARMM27 group charges are shown in parentheses. The two dashed lines indicate transfer schemes A and B (see Sec. 3.2.2 for details). Note that the precision of the charges leaves DPePC with an apparent non-zero net charge due to round-offs. This is not the case when all six decimals are included (see Table 7.1 Appendix 7.1).

and was part of our rationale for using several DPePC configurations.

Even though we do not make use of charge groups in our simulations, we calculated the net charge of the CHARMM27 charge groups for the new charge sets. These are shown in Fig. 3.2 by the grey boxes and the corresponding group charges are denoted q_{X} (X is the group number I-VII). The group charges for CHARMM27 are shown in parentheses. Even though the RESP method predicts that N3 has a positive charge, the direction and magnitude of the NP-dipole moment is similar to that predicted by the Mulliken method. Comparison of $\mathrm{q_I}$ and $\mathrm{q_{II}}$ indicates that the new charges, both Mulliken and RESP, give a smaller NP-dipole moment compared to CHARMM27 charges. Interestingly, the methylene group charges ($\mathrm{q_V}$ and $\mathrm{q_{VI}}$) are no longer neutral and therefore do not have alkane properties as in the CHARMM27 parameter set.

Since there is no rigorous definition of atomic partial charges, any of the three charge distributions in Fig. 3.1 can be valid and we need to test their quality in lipid bilayer simulations. Therefore, the charges found for DPePC were transferred to DPPC using schemes A and B (see Sec. 3.2.2). To obtain a neutral DPPC molecule after the charge transfer we balanced out the net charge with small counter charges on each non-acyl atom in DPPC. These counter charges are numerically smaller than $\sim 10^{-3}$ e, which is less than the standard deviation of the new charges (error bars in Fig. 3.1). Table 7.1 and Fig. 7.1 in Appendix 7.1 provides an overview of our charges after transfer to DPPC.

3.4.2 Testing new partial charges

We tested parameter sets I.A, I.B, II.A and II.B in MD simulations of a DPPC lipid bilayer. For reference we have also included results from two simulations using the CHARMM27 force field, one with no applied surface tension and one with an applied surface tension of 61 mN/m [82]. These reference simulations are referred to as III and IV, respectively. All six simulations are summarized in Table 3.1. In the following we compare the area

Sim. ID	Parameters	γ (mN/m)	length (ns)
I.A	New Mulliken charges assigned down to the 3rd acyl carbon	0	2
I.B	New Mulliken charges assigned down to the 2nd acyl carbon	0	2
II.A	New RESP charges assigned down to the 3rd acyl carbon	0	17
II.B	New RESP charges assigned down to the 2nd acyl carbon	0	2
III	CHARMM27	0	15
IV	CHARMM27	61	10

Table 3.1: Overview of MD simulations. Simulations I.A, I.B, II.A, II.B use new atomic partial charges, whereas simulations III and IV use the CHARMM27 partial charges. The simulation lengths are shown in the rightmost column. Simulations with $\gamma = 0$ mN/m correspond to the NPT ensemble with $P_x = P_y = P_z$ =1 bar.

per lipid, the volume per lipid, the order parameter profile and the electron

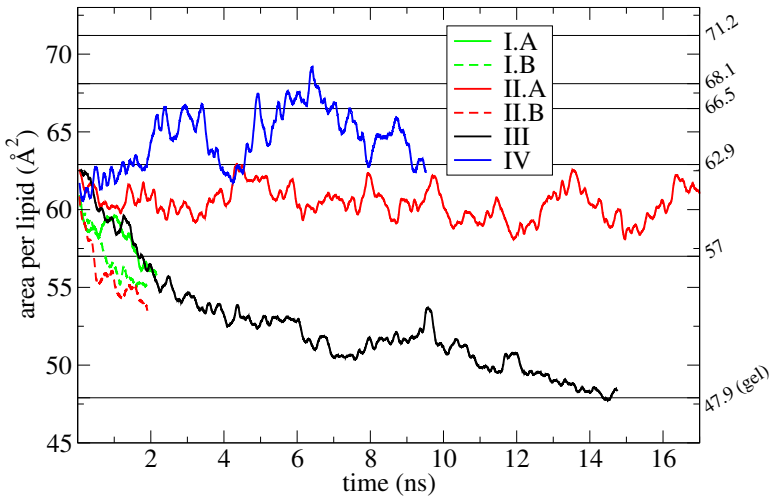


Figure 3.3: Area per lipid as function of time for six simulations using different parameter sets and ensembles (see Table 3.1). The curves are 200 point running averages of the area per lipid calculated every 0.5 ps. The numbers on the right ordinate axis are experimental values for the area per lipid [116].

density profile for these six simulations mutually and with experiments.

Area and volume per lipid

From Fig. 3.3 we see that the areas per lipid in simulations I.A, I.B and II.B monotonically decrease to ∼55 Å^2 within the first 2 ns of the simulations, resembling the behavior of the first 2 ns of simulation III. In simulation III, the area per lipid reaches the experimental value for the gel phase [116] after 15 ns and is still decreasing moderately. Based on the immediate decrease in the area per lipid observed in simulations I.A, I.B and II.B we decided to end these three simulations after 2 ns and to discard them from further analysis. In contrast, the area per lipid found in simulation II.A is stable and the average value of 60.4±0.1 Å^2 compares favorably to the commonly accepted experimental value of 64 Å^2 [116]. As previously reported, it is necessary to apply a surface tension of 61 mN/m in order to obtain an area per lipid of 64.5±0.3 Å^2 using the CHARMM27 force field (simulation IV) [82]. The error estimates for the area per lipid are obtained as the standard error of the mean area calculated in 250 ps data blocks.

Simulations II.A, III and IV give lipid volumes of (12.0±0.3) · 10^2, (11.7±0.6) · 10^2 and, (12±1) · 10^2 Å^3/lipid. The experimental estimate is ∼ 12.30 · 10^2 Å^3/lipid and varies a few Å^3/lipid depending on the technique used [116]. Thus, all the simulation results agree with the experimental data within the statistical uncertainty in the simulations. In the determination of the lipid volume, the water volume was calculated from the water electron density in Fig. 3.5(a). Simulation III is not completely equilibrated

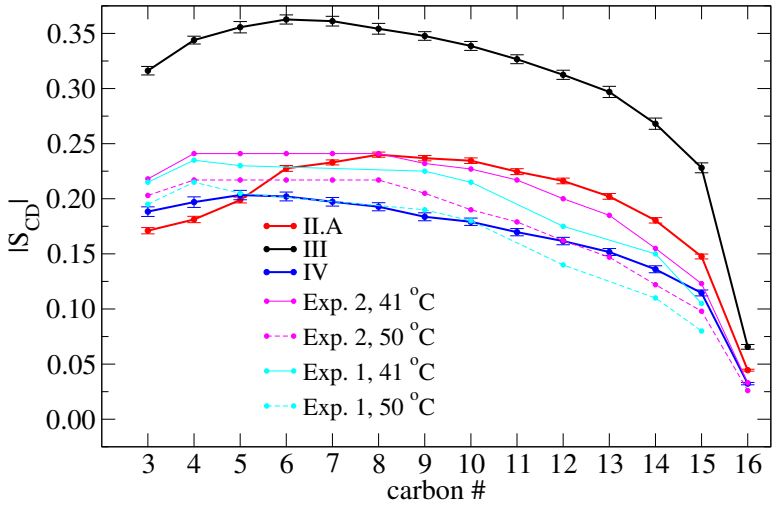


Figure 3.4: The order parameter profiles from two experiments. (Exp. 1 and Exp. 2) and three simulations (II.A, III and IV) for C_3 to C_{16}. Exp. 1 refers to the experiments of Seelig and Seelig [132], while Exp. 2 refers to the experiments of Douliez et al. [48]. For simulation III, the order parameters are calculated from the last 3 ns. The error bars represent the standard error of the mean and were estimated from the variations in the average order parameter profiles calculated in 250 ps time blocks.

(cf. Fig. 3.3) and therefore we used only the last 3 ns to estimate the lipid volume in this simulation. In the following sections we also use only the last 3 ns of this simulation for the data analysis.

Order parameters

In Fig. 3.4 we show the order parameter profile for the sn2-chain from simulations II.A, III and IV as well as experimentally determined deuterium order parameters at 41 °C and 50 °C [48, 132]. The experimental profile measured at 50 °C is close to the simulation conditions. The profile obtained at 41 °C is the upper limit for $|S_{\mathrm{CD}}|$ in a fluid phase DPPC bilayer, since the main phase transition temperature is 41 °C [132]. In the region from C_7 to C_{16}, the profile from simulation II.A resembles most closely the experimental profile for the fluid phase found at 41 °C. In the region nearer to the glycerol backbone (C_3-C_6) the order parameters in simulation II.A are lower than the experimental values at 50 °C and therefore the chains are less ordered than expected. Such deviations are not found in simulation IV (CHARMM27 parameters with $\gamma = 61$ mN/m), which resembles the experimental profiles at 50 °C quite closely. As reported by Feller and MacKerell Jr. [51], the order parameter for the second acyl carbon (C_2) is underestimated in the CHARMM27 force field. Simulation IV confirms this result and the same underestimation is found in simulation II.A (data not shown). The order parameter profile calculated from simulation III (CHARMM27, $\gamma = 0$ mN/m)

clearly indicates highly ordered acyl chains characteristic for the gel phase.

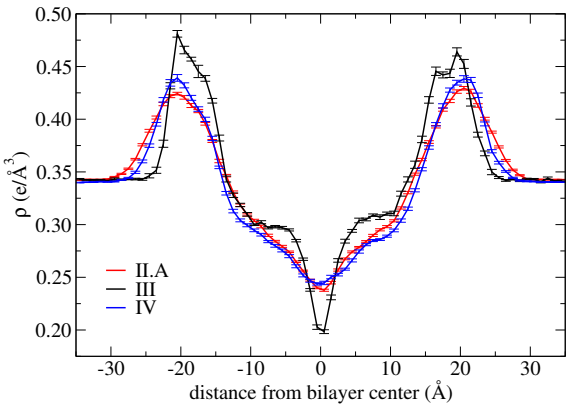
Electron density profile

In Fig. 3.5(a) we show the average electron density profile from simulation II.A, III and IV. The overall shapes of the three profiles are similar and the characteristic methyl troughs in the bilayer center and the headgroup peaks are apparent in all three simulations. The profile from simulation III, has more sharp features than the two other profiles, but the differences between simulations II.A and IV are more subtle. To obtain the bilayer form factors $F^{\mathrm{sim}}(q)$, we Fourier transform the real space profiles in Fig. 3.5(a), which yields the $F^{\mathrm{sim}}(q)$ curves in Fig. 3.5(b) [95]. We have also included estimates of the absolute bilayer form factors from X-ray scattering experiments [117] and fitted these to our simulated $F(q)$. For the fitting, we scaled the experimental first order (h=1) form factors $F^{\mathrm{exp}}(q^{\mathrm{exp}}_{h=1})$ to the simulated form factors $F^{\mathrm{sim}}(q^{\mathrm{exp}}_{h=1})$ and scaled the higher order form factors of the same experimental sample by the same amount. For $h \geq 2$ the experimental $F(q)$ is larger than observed in both simulations. The root mean square deviations of the fitted experimental data points from the corresponding simulated $F(q)$ are 0.4, 1.1 and 0.2 e/Å^2 for simulations II.A, III and IV, respectively. Judged by the form factors, simulation IV therefore exhibits the best agreement with the experimental data. Figure 3.5(c) shows new and improved X-ray results from the laboratory of John F. Nagle.

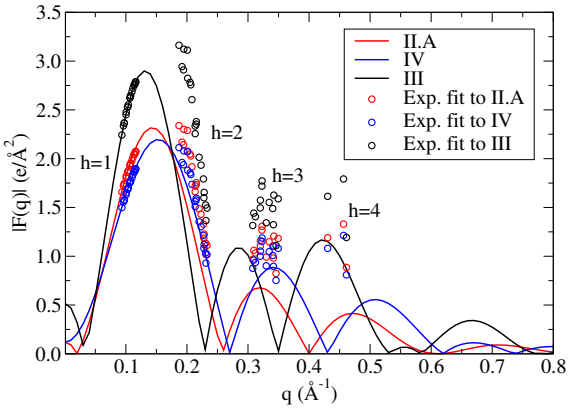
3.5 Discussion

MD simulations of DPPC lipid bilayers using the all-atom CHARMM27 parameter set yield highly ordered bilayers with gel-like properties in the NPT ensemble [53, 82]. Since finite size effects only account for a part of the bilayer ordering, it appears that the CHARMM27 lipid parameters need optimization provided that one resorts to NPT simulations.

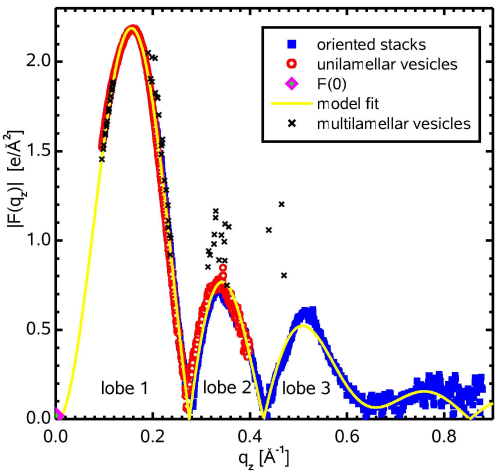
In this study we determined four new sets of partial charges for the DPPC lipid headgroup from Mulliken population analysis and from the RESP fitting procedure. These procedures do not comply with the traditional strategy used for optimizing partial charges in the CHARMM force fields [51, 56, 107, 108, 130]. To this end we note that Foloppe and MacKerell, Jr. [56] explains that the CHARMM27 parameters are "primarily optimized to reproduce macromolecular target data while maximizing agreement with small molecule target data". In our opinion, the macroscopic target data are the key properties of the bilayer and consequently a correct representation of these must be the ultimate goal of any simulation deploying CHARMM lipid parameters. Thus, even though the methods we used to obtain the new atomic partial charges deviate from the traditional CHARMM parameterization strategy, this should be of minor importance if the new



(a)



(b)



(c)

Figure 3.5: Electron density data. (a) Electron density profiles from simulations II.A, III and IV. The error bars represent the standard error of the mean and were estimated from the variations in the average electron density profiles calculated in 250 ps time blocks. (b) Absolute form factors obtained by Fourier transform of (a). We have included estimates of the absolute form factors obtained from experiments [117]. These were scaled such that $F^{\mathrm{exp}}(q^{\mathrm{exp}}_{h=1}) = F^{\mathrm{sim}}(q^{\mathrm{exp}}_{h=1})$ (open circles). (c) New and improved experimental form factors $F(q_z)$ for DPPC at 50 °C (courtesy of John F. Nagle). The black crosses are the old experimental results which were used for the fit in panel (b).

parameters represent DPPC bilayers more accurately than the conventional CHARMM27 parameters.

We tested the four new sets of partial charges in MD simulations using the charges presented in Table 7.1 in Appendix 7.1, i.e., with six decimals. Using atomic partial charges with six decimals in the MD simulations should not be crucial for the properties of the bilayer. However, keeping this precision was convenient since the round-off errors introduced to fulfill electroneutrality were found smaller than the configurational dependence of the atomic charges.

3.5.1 Which parameter set?

Figure 3.3 shows that within 15 ns, the area per lipid in simulation III decreases to $\sim$48 Å^2, which is close to the experimental value for the DPPC gel phase [116]. Based on the similarity of the first few nanoseconds, the same behavior is expected for simulations I.A, I.B and II.B. Thus, further equilibration of these simulations would most likely give bilayers with gel-like properties. Since the goal of this study is to develop a new parameter set that gives more fluid-like bilayers in the NPT ensemble, we stopped simulations I.A, I.B and II.B after 2 ns. The area per lipid in simulation II.A oscillates around an average value of 60.4 Å^2 and considering the variations in the area per lipid as determined from different experiments [116], this value is satisfactory. The area per lipid from simulation II.A is also in excellent agreement with the results of MD simulations using other force fields [4, 19, 158]. In order to obtain an area per lipid that is close to the experimental value of 64 Å^2 using the CHARMM27 parameters, it is necessary to apply a surface tension of 61 mN/m (simulation IV) [82]. The volume per lipid is comparable to the experimental value of 1230 Å^3 [116] for all three simulations.

The acyl chains in simulation III were found to be highly ordered when compared to the experimental data for a fluid (L_α) phase bilayer, which is consistent with the underestimation of the area per lipid in this simulation. The chain order data are similar to results from bilayer simulations using the CHARMM22 parameter set with zero applied surface tension [53]. The order parameter profiles for simulations II.A and IV indicate that these simulations are in the fluid phase. In simulation II.A, the lower part of the acyl chains is slightly more ordered than expected from experiments, but is still fluid-like when compared to simulation III. In the region C_3-C_5 the chain order is underestimated and although this underestimation is small compared to the overestimation in simulation III, it is evident that the II.A order parameter profile is not practically constant from C_3 to C_{8-11} as found in experiments [48]. Simulation IV resembles the experimental profiles at 50 °C and reproduces the plateau from C_3 to C_{8-11}. This indicates that the non-constant order parameter profile from C_3 to C_{8-11} in simulation II.A

is an artifact of the II.A parameters and is not a consequence of melting of the acyl chains. Since we assigned new charges to the first four carbons of the acyl chains (C_1-C_4) in the II.A parameter set without readjusting the dihedral parameters accordingly, it seems likely that the new parameters have lowered the gauche energy in the C-C-C-C torsional potential in the upper acyl chain region. However, this is not the case, since we find that the torsional potential shows a slight increase in the gauche energy for rotation around the C_2-C_3 and C_3-C_4 bonds of the pentanoate chain in ethylpentanoate with II.A parameters (data not shown). Seemingly, the unscreened charge of the fourth acyl carbons (C_4) engage in intra and/or inter lipid electrostatic interactions that are not accounted for in the torsional energy surface of the ethylpentanoate.

Compared to simulations II.A and IV, the electron density profile from simulation III shows sharp head group peaks, a deep narrow methyl trough in the bilayer center and plateau regions in between. These well-defined features indicate that the positions of the head group phosphate moiety and the terminal methyl groups are relatively well defined, which is consistent with the overall picture of a very ordered bilayer structure in this simulation. The sharp features in the electron density profile of simulation III are also reflected in the form factor $F^{\mathrm{III}}(q)$, which is non-zero for $q > 1$ Å^{-1} and resembles the experimental gel phase $F(q)$ [116]. However, the bilayer in simulation III is not in a fully developed gel ($L_{\beta'}$) phase since the acyl chains are not tilted relative to the bilayer normal. The difference between the electron density profiles from simulations II.A and IV are subtle and these are compared with experimental data in Fourier space. Fitting the experimental estimate for the absolute form factors to the simulated factors we find that the experimental data points consistently lie slightly higher than predicted by both simulations II.B and IV which, except for the third order data, is consistent with the findings of Sachs et al. [128]. Overall, the comparison of absolute form factors indicates that simulation IV conforms more closely to X-ray scattering experiments [117] than the results from simulation II.A. For this comparison we used the currently available experimental data, however, the new and improved X-ray results for DPPC shown in Fig. 3.5(c) agree much better with both simulations II.A and IV than the data used for comparison in Fig. 3.5(b). For example, it appears that the experimental form factors we have used for the fitting in Fig. 3.5(b) are overestimated by ~ 0.3 e/Å^2 and ~ 0.5 e/Å^2 for $h = 3$ and $h = 4$, respectively, compared to the new data in Fig. 3.5(c). Until the new DPPC X-ray data is published, we cannot carry out a more thorough comparison to determine which one of simulations II.A or IV reproduce the new experimental data the best. The new experimental X-ray data is obtained by the methods previously used for dilauroylphosphatidylcholine (DLPC) and dimyristoylphosphatidylcholine (DMPC) bilayers [96].

The II.A parameters were previously used in a pressure profile study [137].

The pressure profiles reported in that study qualitatively resemble pressure profiles calculated from other atomistic force fields [62, 98]. Since pressure profiles cannot be measured experimentally we will settle with this qualitative agreement.

In summary, our investigations of the area and volume per lipid, the order parameter profile and the electron density profile for DPPC clearly indicate that, in the NPT ensemble, the II.A parameters reproduce fluid phase bilayer properties better than the CHARMM27 parameters. The volume per lipid compares favorably with experiments for simulations II.A, III and IV. Further we find that for a pure DPPC lipid bilayer simulated with CHARMM27 charges, excellent fluid-like properties can also be obtained by applying an appropriate positive surface tension which, however, requires that the area per lipid has been pre-determined experimentally.

3.5.2 The origin of the fluid phase properties.

For future optimizations studies of the CHARMM lipid parameters, it is useful to pin down why the II.A parameters improve the bilayer properties. It is likely that simulations I.A, I.B and II.B would eventually attain gel-like properties as does simulation III, which indicates that area per lipid is relatively insensitive to most of the changes that we have made in the head group, such as the inversion of the sign on the N3 and C6 charges in parameter set II.B. Apparently, the reduction of the group charge of the choline and phosphate moieties (reduction of the NP-dipole moment) in simulations I.A, I.B and II.B does not affect the area per lipid either. Essentially the only difference between simulations II.A and II.B is the partial charge of the C_4 methylene group. Since only simulation II.A is able to maintain fluid phase properties of the bilayer, the C_4 methylene group seems to be responsible for the fluid phase properties of the DPPC bilayer in that simulation.

This extreme and unexpected sensitivity to the charges in the C_4 methylene group stresses the importance of fine tuning the lipid parameters in the glycerol backbone region. Since the glycerol backbone is a common motif in all glycero phospholipids such a fine tuning could be useful also for other lipid systems.

3.6 Summary and Conclusion

As previously described in the literature, we confirm that dipalmitoylphosphatidylcholine (DPPC) lipid bilayer simulated in the NPT ensemble using the CHARMM27 force field have gel-like properties and that excellent fluid (L_α) phase properties can be obtained by applying an appropriate positive surface tension [51, 53, 82, 94]. Since the applied surface tension has to be adjusted based on experimental data [53, 82], we believe that this approach

makes it difficult to take full advantage of the predictive power of MD simulations. In an attempt to obtain fluid phase properties of DPPC bilayers simulated in the NPT ensemble using the CHARMM energy function, we assigned new partial charges to the head group and upper acyl chains of DPPC using Mulliken population analysis and a RESP fitting procedure. The new Mulliken partial charges do not have an immediate effect on the bilayer properties, but the RESP charges do. Using the new RESP charges, we find a dramatic improvement of the bilayer properties compared to simulations using the CHARMM27 charges with zero applied surface tension. Thus, the new RESP partial charges presented in this study for the first time allow for simulating a DPPC lipid bilayer in the fluid phase at constant pressure and zero applied surface tension using the CHARMM energy function.

4 Pressure profile calculations in lipid bilayers

Based on: J. Sonne, F.Y. Hansen, and G.H. Peters. Methodological problems in pressure profile calculations for lipid bilayers. Chem. Phys., **122** 124903 (2005).

4.1 Introduction

A system consisting of two or more phases will have interfacial regions that separate the bulk phases. The properties of interfacial regions can be characterized through macroscopic variables such as the surface tension, surface free energy, bending modulus, saddle splay modulus etc.. These properties can be determined from the pressure distribution through the interfacial region [127, 129, 140] (the pressure profile). Therefore, pressure profiles provide a microscopic interpretation of interfacial phenomena allowing for instance determination of the viscosity profile [2, 80], which is of great interest in materials and polymer science.

Pressure profiles are not available experimentally [111], but can be calculated from e.g. molecular dynamics (MD) simulations (see e.g. Refs. [1, 58, 61, 98, 157]). Early MD studies of simple liquid-gas interfaces [1, 157], explored methods for calculating the local pressure and provided insight into the microscopic properties of the surface tension. In biological systems, such as cell membranes, the pressure profile plays a central role since the function and survival of living cells is tightly coupled to the mechanical properties of the membrane [21]. Moreover, it has been suggested that the pressure profile in cell membranes undergoes changes in the presence of anesthetic compounds and that this change might be a key step in general anesthesia [29, 31]. Changes in the pressure profile and lateral stress in lipid bilayers, which is a common model system for cell membranes, have also been shown to affect the function of mechanosensitive channels [38, 61]. Computer simulations [58, 61, 62, 98] of lipid bilayers indicate that bilayers have regions with negative lateral pressure trying to minimize the interfacial area, and regions with positive lateral pressure trying to expand the bilayer. The mentioned MD bilayer studies, as well as less detailed models [76, 111, 140]. predict lateral pressure variations in these regions of several hundred bar.

In order to determine the pressure profile one needs to calculate the local pressure. The local pressure is not uniquely defined since the expression for the local pressure involves an integral along an arbitrarily chosen contour [131]. In previous studies, two contours have been employed leading to two different expressions for the local pressure tensor, namely the Irving-Kirkwood [74] (IK) and the Harasima [65] (H) local pressure tensors. An obvious dilemma in pressure profile calculations is that it is not possible to rule in favor of either of these expressions. In most bilayer studies the IK expression has been preferred [58, 61, 62, 98].

The IK expression is applicable for interactions described by m-body potentials, where m is finite [58], which makes it possible to include local pressure contributions from e.g. valence angle and dihedral interactions. There is no problem in including electrostatic interactions when these are calculated directly from the Coulomb potential ($m = 2$). For computational reasons, calculating electrostatic interactions directly from the Coulomb potential requires the use of a cutoff and due to the long-range nature of electrostatic interactions, the cutoff should be fairly long. However, even when the cutoff is long, structural artifacts may be introduced in the system [122, 152] and therefore the use of an electrostatic potential cutoff is not always an optimal choice [2]. One way to avoid this is to evaluate the electrostatic interactions by the Ewald summation technique (see e.g. Ref. [2]), in which the electrostatic interactions are split up in a real and reciprocal space contribution. The real space contribution is still pair-wise additive with $m = 2$, while m is formally infinite in the reciprocal space sum. Consequently, it is not possible to include the contribution from the reciprocal space part of the Ewald sum in the pressure profile using the IK expression for the local pressure tensor. One approach [61, 98] to this problem has been to calculate the pressure profile by the IK expression, using a large cutoff for the electrostatic interactions. When the simulations are conducted using a cutoff, the pressure profile should also be calculated using the same cutoff for consistency. However, when the simulations are conducted using Ewald summation while the pressure profile is calculated with a finite cutoff, the long-range electrostatic interactions are neglected in the latter calculation. The neglected contribution to the lateral pressure may be relatively small, but one should keep in mind, that the local lateral pressure may vary several hundred bar through the bilayer even though bulk pressure in the simulation is merely 1 bar [58, 61, 62, 98]. Thus, a small contribution to the total simulation pressure might not be negligible in pressure profile calculations. A method for including the reciprocal space contribution was developed by Alejandre et al. [1] in a study of a water-vapor interface. In this method, they used different expressions for the real space and reciprocal space electrostatic contributions to the pressure profile and their method was only set up to handle systems with one kind of molecules that were assumed to be rigid.

The present chapter addresses two central problems in pressure profile calculations. The first problem is the arbitrariness in the choice of integration contour. We will investigate this problem by comparing the pressure profiles obtained with the two different contours that lead to the IK and H expressions for the local pressure. The second problem is how to include the electrostatic contribution to the local pressure tensor when Ewald summation is used in a multi-component system with flexible molecules.

The chapter is organized as follows: First, in Sec. 4.2, we summarize the theory of local pressure calculation and show how the H expression can be used to calculate the local pressure contribution from the Ewald sum evaluation of electrostatic interactions in systems with different and flexible molecular species. In Sec. 4.4.1 the differences between pressure profiles calculated according to the IK and H expressions are investigated for a dipalmitoyl-phosphatidyl-choline (DPPC) lipid bilayer. Based on these investigations, we discuss the uniqueness of the pressure profile for the system in consideration. In Sec. 4.4.2, the pressure profile including all electrostatic interactions as determined by the Ewald summation technique, is compared to pressure profiles calculated with varying cutoffs for the electrostatic interactions. From this comparison, we discuss the long-range electrostatic contribution to the pressure profile. The results in Sec. 4.4.2 is followed by a summary and the conclusion. In Sec. 4.3 we describe the system setup and other simulation details.

4.2 Theory

4.2.1 Bulk pressure tensor

The bulk pressure tensor for a system is a sum of kinetic and configurational contributions i.e. $\mathsf{P}^{\text{tot}} = \mathsf{P}^{\text{kin}} + \mathsf{P}^{\text{config}}$. In this chapter, only the configurational part of the pressure tensor is of interest and for simplicity it will be denoted P. The $\alpha\beta$-element of the configurational part of the pressure tensor is obtained from the virial theorem [2]

$$\mathrm{P}^{\alpha\beta} = \left\langle \frac{1}{V} \sum_i f_i^\alpha r_i^\beta \right\rangle = \left\langle -\frac{1}{2V} \sum_{i \neq j} f_{ij}^\alpha r_{ij}^\beta \right\rangle, \tag{4.1}$$

where V is the volume of the system, $\mathbf{f}_i$ the total force on particle i and $\mathbf{r}_i$ its position. The brackets denote time averaging. In the second part of the equation we have used $\mathbf{r}_{ij} = \mathbf{r}_j - \mathbf{r}_i$ and have assumed that the interactions in the system may be written as a sum over pair-interactions, that is $\mathbf{f}_i = \sum_j' \mathbf{f}_{ij}$ where $\mathbf{f}_{ij}$ is the force on particle i due to particle j. The prime in the summation indicates that $j \neq i$.

4.2.2 Local pressure tensor

By setting up a microscopic momentum balance and relating this to the corresponding continuum expression one arrives at the following expression for the configurational part of the local pressure tensor [131, 156, 157]:

$$\mathrm{P}^{\alpha\beta}(\mathbf{r}) = \left\langle \sum_i f_i^\alpha \int_{\mathcal{C}_{0i}} \delta[\mathbf{r}-\mathbf{l}]\mathrm{d}s^\beta \right\rangle . \tag{4.2}$$

The contour integral runs along an arbitrary path $\mathcal{C}_{0i}$ from a reference position $\mathbf{R}_0$ to the position of the i'th particle $\mathbf{r}_i$. $\mathbf{s}$ is a line element on $\mathcal{C}_{0i}$ and $\mathbf{l}$ is the position vector of the line element. $\delta[\mathbf{r}-\mathbf{l}]$ is the Dirac delta function. Equation (4.2) expresses that the pressure tensor near the point $\mathbf{r}$ is a sum of contributions coming from all particles $\{k\}$ for which the corresponding contours $\{\mathcal{C}_{0k}\}$ pass through the region around $\mathbf{r}$.

For a pair-wise additive potential, Eq. (4.2) can be expressed as [127, 131, 156, 157]:

$$\mathrm{P}^{\alpha\beta}(\mathbf{r}) = \left\langle -\frac{1}{2}\sum_{i\neq j} f_{ij}^\alpha \int_{\mathcal{C}_{ij}} \delta[\mathbf{r}-\mathbf{l}]\mathrm{d}s^\beta \right\rangle . \tag{4.3}$$

In the derivation of Eq. (4.3) from Eq. (4.2), the contour $\mathcal{C}_{0i}$ is forced through $\mathbf{r}_j$ and the integral is split into an integral over $\mathcal{C}_{0j}$ minus an integral over $\mathcal{C}_{ij}$. The integral along $\mathcal{C}_{0j}$ is zero no matter how the contour is chosen due to the translational invariance of the potential, while the integral along $\mathcal{C}_{ij}$ in general depends on the path chosen [58].

In the following we present the local pressure expressions based on two different choices of $\mathcal{C}_{ij}$. The expressions for the local pressure will be written for a system of planar geometry and will be discretized to slabs of finite size which makes the expressions suitable for implementation in an MD program. Note that the bulk pressure tensor in Eq. (4.1) is a simple average of the local pressures in all slabs. This is used as a consistency check of the calculations.

Irving-Kirkwood contour

The IK expression for the local pressure tensor [74] is obtained by choosing the contour $\mathcal{C}_{ij}$ as a straight line connecting the particle pair ij [127, 131, 156, 157]. This contour (IK-contour) is illustrated in Fig. 4.1 (open arrows). In planar geometry (rotational symmetry around the z-axis) the configurational contribution to the lateral pressure, $\mathrm{P_L} \equiv (\mathrm{P}^{xx}+\mathrm{P}^{yy})/2$, in slab s, extending from z_{sl} to z_{su} with volume V_s, is given by

$$\mathrm{P_L}(s) = \left\langle -\frac{1}{2V_s}\sum_{i\neq j} \frac{f_{ij}^x r_{ij}^x + f_{ij}^y r_{ij}^y}{2} F_s(z_{su}, z_{sl}, r_i^z, r_j^z) \right\rangle, \tag{4.4}$$

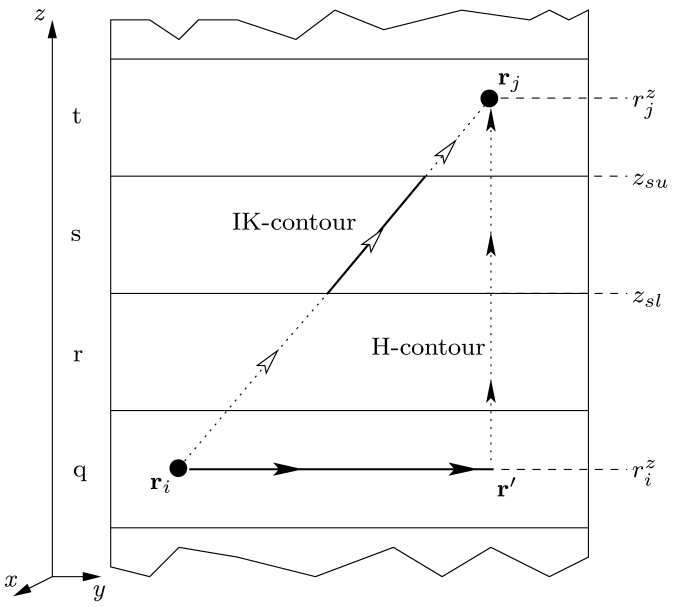


Figure 4.1: Illustration of the contours leading to the IK and H expressions for the local pressure tensor (open and filled arrows, respectively). The bold part of the IK-contour indicates the fraction of the full virial from the interaction between i and j that is assigned to slab s according to Eq. (4.4). The bold part of the H-contour indicates that the virial of the interaction from i to j is assigned to slab q according to Eq. (4.6).

where F_s is the fraction of the connecting line between i and j that is located within slab s i.e. between z_{sl} and z_{su}. In Fig. 4.1, F_s is given by the ratio between the length of the bold part of the IK-contour and the length of the full IK-contour which can be expressed as

$$F_s = |z_{su} - z_{sl}|/|r_j^z - r_i^z|. \tag{4.5}$$

The expression in Eq. (4.4) can be generalized to m-body interactions with m finite [58] and it is therefore straight-forward to include local pressure contributions from valence angles ($m = 3$) and dihedral interactions ($m = 4$). Electrostatic interactions can also be incorporated using Eq. (4.4) if all electrostatic interactions are evaluated directly from the Coulomb potential. However, using the Ewald summation technique [2] introduces a problem in the pressure profile calculation since the reciprocal space part of the sum can not be expressed as m-body interactions with m finite and it is therefore not possible to determine F_s in Eq. (4.5).

Harasima contour

The contour leading to the Harasima expression for the local pressure tensor [65] goes from particle i at $\mathbf{r}_i = r_i^x\hat{\mathbf{e}}_x + r_i^y\hat{\mathbf{e}}_y + r_i^z\hat{\mathbf{e}}_z$ to particle j at $\mathbf{r}_j = r_j^x\hat{\mathbf{e}}_x + r_j^y\hat{\mathbf{e}}_y + r_j^z\hat{\mathbf{e}}_z$ via straight lines through the intermediate

point $\mathbf{r}' = r_j^x \hat{\mathbf{e}}_x + r_j^y \hat{\mathbf{e}}_y + r_i^z \hat{\mathbf{e}}_z$ (H-contour, lines marked with filled arrows in Fig. 4.1). The resulting expression for the lateral pressure in slab s is [65, 131, 156, 157]

$$\mathrm{P_L}(s) = \left\langle -\frac{1}{2V_s} \sum_{i \neq j} \frac{f_{ij}^x r_{ij}^x + f_{ij}^y r_{ij}^y}{2} \Theta\left(z_{su} - r_i^z\right) \Theta\left(r_i^z - z_{sl}\right) \right\rangle, \qquad (4.6)$$

where $\Theta(x)$ is the Heaviside step function. Equation (4.6) expresses that half of the lateral pressure arising from the interaction between i and j is assigned to the slab where i is located (cf. bold part of the H-contour in Fig. 4.1). The other half of the lateral pressure from this pair is assigned to the slab where j is located.

Properties of the IK and H expressions

At this point we have two expressions for the local pressure and when m is finite both Eqs. (4.4) and (4.6) can be applied with almost equal simplicity. However, there are two important differences between the expressions that will be discussed in the following:

(1) The IK expression distributes the virial evenly on the connecting line between two interacting particles and therefore the contribution to the local pressure from this pair is the same in all slabs in the region between the two particles. Further, the local pressure is independent of the number of slabs when the slab width is smaller than the distance between the two particles ("range of interaction") except for the slabs with the particles. There, the connecting line between the particles does not cross the entire slab. If the lateral contribution to the virial from this interaction is W_{12}, then the lateral pressures in all seven slabs are the same and equal to $(\frac{1}{7}W_{12})/(Ah)$, where A is the xy-area of the simulation box and h is the slab width. If we were using only one slab of height $7h$ (as indicated to the left in Fig. 4.2(a)) the lateral pressure in this slab would also be $(W_{12})/(7Ah)$. To the right in Fig. 4.2(a) the corresponding pressure distributions are shown and the IK-1 and the IK-7 profiles (IK-profile using 1 and 7 slabs, respectively) coincide. This property can also be seen from Eq. (4.4) where the slab volume V_s and the fraction of connecting line in the slab F_s both scale linearly in the slab width. Thus, the pressure in a slab is independent of the slab width. These simple considerations also indicate that, for one pair, the IK expression will not predict local pressure variations on length-scales that are shorter than the relevant "range of interaction". Therefore, it is reasonable to use a slab width which is comparable to the shortest "range of interaction" in the system.

The local pressure distribution from the H expression is also shown in Fig. 4.2(a). With one slab, the lateral pressure distribution, is identical to the IK-profiles as illustrated by the H-1 profile. This illustrates that

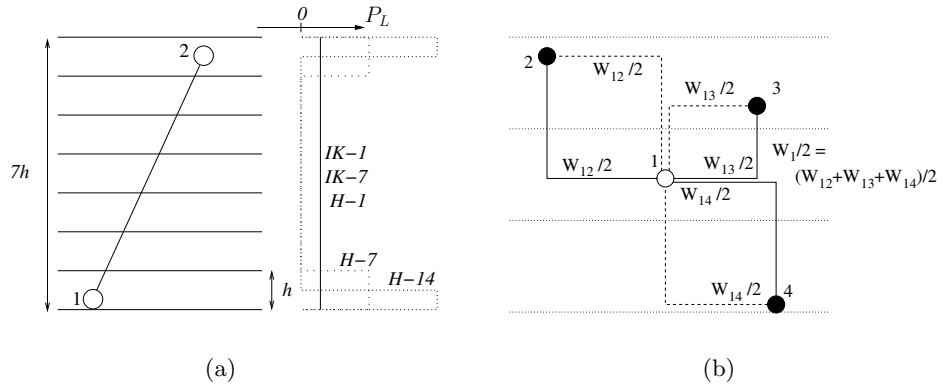


Figure 4.2: Properties of the IK and H methods for evaluating the pressure profile (a) Schematic illustration of the local pressure distribution coming from the IK and H expressions (solid and dotted line, respectively) when varying the slab width. See text for details. (b) Four particle system. The virial contributions from particle 1 (half of W_{12}, W_{13} and W_{14}) all contribute to the local pressure in the slab where particle 1 is located. The sum of these contributions is identical the half of the total virial associated with particle 1 (W_1). Likewise, virial contributions from particle 2 (e.g. $W_{21} = W_{12}$) contribute to the local pressure in the slab where particle 2 is located, however, this is in principal not important when considering particle 1.

the local pressure is independent of the choice of contour and therefore well-defined on the length-scale of the "range of interaction". This is true for any contour that is confined to the region between the particles. When the slab width is decreased to h, i.e. we are using seven slabs, the H expression gives the pressure $(\frac{1}{2}W_{12})/(Ah)$ in the two slabs where the particles are located and zero elsewhere (H-7 profile in Fig. 4.2(a)). Increasing the number of slabs (thereby reducing the slab width) will cause the H-pressure, in the two slabs with the particles, to increase which is illustrated by the H-14 profile in Fig. 4.2(a) (14 slabs between the particles). The pressure peaks in the H-profile can be smeared out by the motions of the two particles and for a homogeneous particle density, the IK and H profiles are identical [131]. The invariance of the local pressure to the slab width favors the use of the IK expression over the H expression, but does not imply that IK-profiles are more correct.

(2) As described in Sec. 4.2.2, the IK expression cannot be used to calculate the local pressure contribution from the reciprocal space part of the Ewald sum. This is, however, possible with the H expression: If the virial from the interaction from particle i to particle j is known, one need not know r_j^z in order to use the H expression in Eq. (4.6). Although the reciprocal space sums are not expressed in terms of pair interactions, it is possible to derive an expression for the pressure contribution coming from interactions between particle i and all particles represented in the reciprocal space

part of the Ewald sum (see the following section). This contribution can be considered as sum of many pair-wise additive contributions (between i and different j's) which should all be assigned to the slab where i is located [cf. Fig. 4.2(b)], and therefore the sum of these pair-contributions can be assigned to the slab where i is located. This approach will be referred to as the Harasima-Ewald-method.

Since the vertical part of the H-contour (cf. Fig. 4.1) passes through all slabs between the interacting particles, like the IK-contour, the zz-element of the H pressure tensor depends on two particle positions in a similar way as the IK expression in Eq. (4.4) does. The expression for the zz-element of the H local pressure tensor is obtained by substituting the superscripts x and y with z on f and r in Eq. (4.4) [131, 156, 157]. Therefore, it is not possible to calculate this component of the local pressure tensor from electrostatic interactions evaluated by Ewald summation using the HE method. This, however, does not present a problem since the zz-element of the local pressure tensor (including all interactions) is equal to the bulk pressure everywhere in the simulation box for a system in mechanical equilibrium ($\boldsymbol{\nabla} \cdot \mathsf{P} = \mathbf{0}$). The fact that the H expression allows the reciprocal space part of the Ewald sum to be included in the lateral pressure profile favors the H expression over the IK expression.

4.2.3 Ewald summation and local pressure

In the previous section we argued that the H expression in Eq. (4.6) can be used for distributing the lateral pressure contribution from the reciprocal space part of the Ewald sum, if the pressure contribution from the interaction between particle i and all particles represented in the reciprocal space sums $\mathsf{P}_{K,i}$ is known. Such an expression is derived in the following.

Using periodic boundary conditions, the Coulomb energy U_C of N particles with partial charges $\{q\}$ and positions $\{\mathbf{r}\}$ can be expressed by the Ewald sum [2]:

$$\begin{aligned} U_C =& \frac{1}{2V\epsilon_0} \sum_{\mathbf{k}_n \neq \mathbf{0}} Q(k_n) S(\mathbf{k}_n) S(-\mathbf{k}_n) \\ &+ \frac{1}{4\pi\epsilon_0} \sum_{i<j\in I} q_i q_j \mathrm{erfc}(\kappa r_{ij})/r_{ij} \\ &- \frac{1}{4\pi\epsilon_0} \sum_{i<j\notin I} q_i q_j \mathrm{erf}(\kappa r_{ij})/r_{ij} \\ &- \frac{\kappa}{4\pi^{3/2}\epsilon_0} \sum_i q_i^2. \end{aligned} \tag{4.7}$$

ϵ_0 is the vacuum permittivity, r_{ij} is the distance between particle i and j, κ is a parameter that shifts the energy contribution from the real to the reciprocal space as κ increases. I is the list of particle pairs for which the energy

is evaluated. erf is the error function and erfc is the complementary error function. V is the volume of the simulation box which, for an orthogonal box, is given by $V = L^x L^y L^z$ where L^α is the box length in the α-direction ($\alpha = x, y, z$). $\mathbf{k}_n$ is the reciprocal lattice vector given by

$$\mathbf{k}_n = 2\pi(n^x/L^x, n^y/L^y, n^z/L^z) \begin{pmatrix} \hat{\mathbf{e}}_x \\ \hat{\mathbf{e}}_y \\ \hat{\mathbf{e}}_z \end{pmatrix}, \tag{4.8}$$

with $n^\alpha = 0, \pm 1, \pm 2, \pm 3....$ $S(\mathbf{k}_n)$ is given by

$$S(\mathbf{k}_n) = \sum_{j=1}^{N} q_j \exp(\mathrm{i}\mathbf{k}_n \cdot \mathbf{r}_j), \tag{4.9}$$

with $\mathrm{i}^2 = -1$. $Q(k_n)$ is given by

$$Q(k_n) = \exp(-k_n^2/4\kappa^2)/k_n^2. \tag{4.10}$$

Note that Eq. (4.7) assumes that κ has been chosen so that the real space part of the energy has converged within the central simulation box. In practice κ is chosen such that the real space energy converges within a specified distance that is smaller than half the length of the simulation box. The first term in Eq. (4.7) is the reciprocal space contribution U_K to the total Coulomb energy U_C From Eq. (4.7) one can extract the energy $U_{K,i}$ which represents the energy of particle i due to all other particles represented in the reciprocal space sum. When this energy satisfies $U_K = \frac{1}{2}\sum_i U_{K,i}$ it is given by [68]:

$$U_{K,i} = \frac{q_i}{V\epsilon_0} \sum_{\mathbf{k}_n \neq \mathbf{0}} Q(k_n) \mathrm{Re}\left[\exp(-\mathrm{i}\mathbf{k}_n \cdot \mathbf{r}_i) S(\mathbf{k}_n)\right]. \tag{4.11}$$

Re denotes the real part of the argument. For a rectangular simulation box, the diagonal elements of the instantaneous configurational contribution to the pressure tensor are given by [25, 119] $\mathrm{P}^{\alpha\alpha}V = -(\partial U/\partial L^\alpha)L^\alpha$. Using the energy expression in Eq. (4.11) gives the following expression for the instantaneous reciprocal space pressure contribution from particle i (see Appendix 7.2).

$$\begin{aligned} \mathrm{P}_{K,i}^{\alpha\alpha} V = & \frac{q_i}{V\epsilon_0} \sum_{\mathbf{k}_n \neq \mathbf{0}} Q(k_n) \mathrm{Re}\Big[\exp(-\mathrm{i}\mathbf{k}_n \cdot \mathbf{r}_i) S(\mathbf{k}_n)\Big] \\ & \times \left[1 - 2k_n^\alpha k_n^\alpha \left(\frac{1}{k_n^2} + \frac{1}{4\kappa^2}\right)\right], \end{aligned} \tag{4.12}$$

half of which should be assigned to the slab where particle i is located according to Eq. (4.6). Note that choosing $U_K = \sum_i U_{K,i}$ instead of $U_K =$

$\frac{1}{2}\sum_i U_{K,i}$ yields an expression for $\mathrm{P}_{K,i}^{\alpha\beta}$, which deviate by a factor of 1/2 compared to Eq. (4.12). However, in that case the whole of $\mathrm{P}_{K,i}^{\alpha\beta}$ should be assigned to the slab where particle i is located and therefore one obtains the same pressure profile (and bulk pressure).

4.3 Simulation Details

The lipid bilayer in the present study consists of 72 dipalmitoyl-phosphatidylcholine (DPPC) lipids solvated with 2 000 water molecules resulting in a total of approximately 16 000 atoms. The water molecules were placed around the bilayer using the program Solvate (by H. Grubmüller) and subsequently the water layer surrounding the bilayer was cropped to a rectangular, periodic simulation box with average xyz-dimensions of approximately 49 Å×45 Å×68 Å. The water molecules were represented by the TIP3 water model [87]. The system was equilibrated for 15 ns with the area fixed to the experimental value (62.9 Å^2 as suggested by Nagle et al. [117]). Data for the pressure profile calculations were collected from the next 17 ns of propagation in the NPT-ensemble. In all simulations we used parameter set II.A presented in Chapter 3. All simulations were carried out using the MD software NAMD [89] with a time step of 1.0 fs. Both bonded and non-bonded interactions were updated every time step. The simulations were conducted at 325 K and an average isotropic pressure of 1 bar. The pressure was controlled by the Nose-Hoover Langevin barostat [55] with a piston oscillation time of 100 fs and a damping time of 50 fs. The three box lengths of the simulation box were allowed to adjust independently to the relevant pressure components. Electrostatic interactions were evaluated using the Particle Mesh Ewald (PME) method [39, 50] with a grid spacing below 1 Å.

Pressure profile calculations were carried out by post-simulation analysis of the trajectory using a program developed by the authors. The program uses routines from the programs Mindy [89] and CatDCD (developed by the Theoretical and Computational Biophysics Group University of Illinois, Urbana-Champaign). In Appendix 7.3 we have included a brief user guide for the program. The profiles were calculated with 70 slabs corresponding to an approximate slab width of 1 Å. The width was adjusted during the calculation according to the fluctuations in the size of the simulation box. In the following the z-axis is defined as normal to the bilayer and $z = 0$ is defined as the z-component of the bilayer center of mass. The statistical analysis was carried out between data blocks averaged over 250 ps which were found to be uncorrelated.

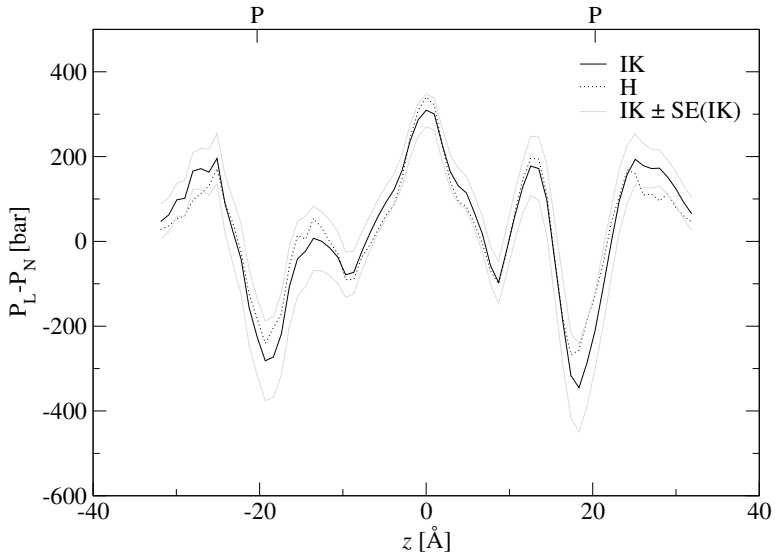


Figure 4.3: Total pressure profiles from the IK and H expressions (4-point running average). Both are calculated with a cutoff of 20 Å for the electrostatic interactions. The z-coordinate is defined normal to the bilayer with origin in the bilayer center of mass. To indicate the dimension of the bilayer, the z-values corresponding to the maximal phosphorus density are marked with the letter P on the top axis of the graph. The solid grey lines show the upper and lower standard deviation [SE(IK)] of the average IK-profile.

4.4 Results and Discussion

4.4.1 Comparing IK and H pressure profiles.

In this section we compare the pressure profiles obtained by the IK and H expressions.

We have shown the total pressure profiles in Fig. 4.3 as calculated by the IK and H expressions with a electrostatic potential cutoff at 20 Å. In the pressure profile calculations the cutoff is limited to 20 Å; a restriction imposed by the box dimensions. Although it is possible to include the reciprocal space contribution in the H-profile, we have chosen to use a finite cutoff to make a direct comparison with the IK-profile possible. We note that both pressure profiles qualitatively resemble the profiles calculated from mean field [13], coarse grained [58], united-atom [98] and all-atom [33, 61] models for lipid bilayers. The small asymmetry in the pressure profiles is probably caused by long wave length membrane undulations with correlation times of more than 5 ns [97, 110]. For completeness we also include Fig. 4.4 that shows all five contributions to the H pressure profile in Fig. 4.3. From Fig. 4.3 we see that the IK and H profiles deviate by less than ±50 bar in most regions and very importantly that the profiles have the same qualitative features. Pressure profiles calculated with electrostatic cutoffs at 8, 12 and 16 Å (not shown) give the same qualitative resemblance between

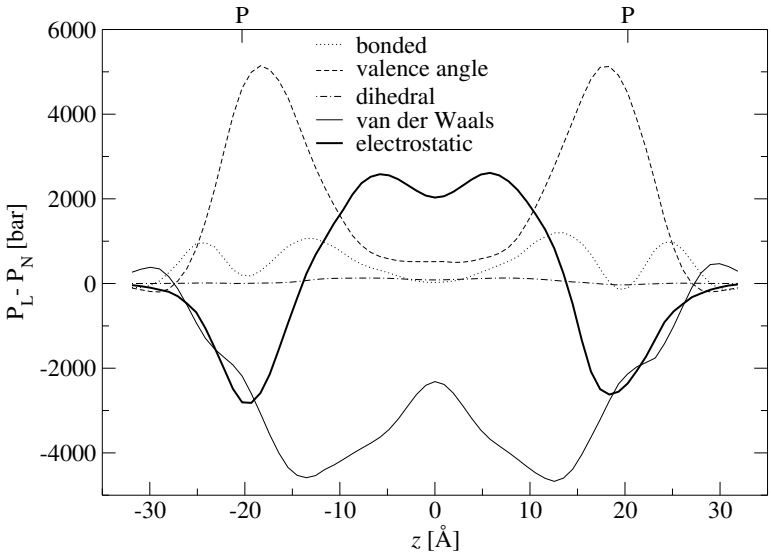


Figure 4.4: Five pressure profile contributions calculated by the H expression for the bonded, valence angle, dihedral, van der Waals and electrostatic interactions in a DPPC lipid bilayer (4-point running average). The sum of these contributions give the total H pressure profile in Fig. 4.3. The z-coordinate is defined normal to the bilayer with origin in the bilayer center of mass. To indicate the dimension of the bilayer, the z-values corresponding to the maximal phosphorus density are marked with the letter P on the top axis of the graph.

the two ways of defining the local pressure. The similarity of the profiles in Fig. 4.3 suggests that the qualitative features of the pressure profiles for this system is rather insensitive to the choice of contour and therefore the pressure profile appears to be well-defined, at least, within a margin of ± 50 bar. The similarity between the IK and H profiles can probably be ascribed to the almost homogeneous particle density throughout the bilayer, with variations of only 10% from the average particle density. Since we only have demonstrated the invariance of the pressure profile to the choice of contour with the IK and H expressions for a DPPC lipid bilayer, we cannot claim the invariance to be universal. For example, in systems with density oscillations, such as liquids near a solid surface [75], qualitative discrepancies between the IK and H pressure profiles have been reported [153].

With that uncertainty in mind, we recommend that the pressure profile in a given system is calculated using both the IK and H expressions to check the uniqueness of the profile. Computing both profiles will not add significantly to the computation time.

Note that if we are only interested in calculating the surface tension γ for an interface ($\gamma = \int_{-L^z/2}^{L^z/2}(\mathrm{P_N} - \mathrm{P_L})dz$), the IK and H expressions will always give the same result [131, 157]. Therefore, irrespective of the nature of the system, the surface tension can be rigorously estimated from the pressure

profile.

4.4.2 Cutoff vs. Ewald summation.

In the previous section we concluded that the pressure profiles obtained from the IK and H expressions are qualitatively similar using cutoffs ranging from 8 to 20 Å. When the Ewald summation technique is used, the energy can no longer be expressed in terms of m-body interactions with m finite as discussed in Sec. 4.2.2, and the electrostatic contributions to the lateral pressure can therefore only be obtained by the H expression (using the HE method). As long as the IK and H expressions lead to the same pressure profile, the HE method allows us to make a unique determination of the pressure profile when the Ewald summation technique has been used in the simulations. In the following pressure profiles calculated by the HE method will be referred to as Ewald-profiles.

As for the electrostatic energy calculated by the Ewald summation technique, the pressure profile should be independent of κ i.e. independent of how the calculation of the profile is partitioned between the real and reciprocal spaces. Thus, we compared two Ewald-profiles calculated with $\kappa = 0.14$ Å^{-1} and $\kappa = 0.40$ Å^{-1} (data not shown). At $\kappa = 0.14$ Å^{-1}, the reciprocal space contribution to the total electrostatic pressure is very small, whereas it accounts for 9% at $\kappa = 0.40$ Å^{-1}. Still, the $\kappa = 0.14$ Å^{-1} and $\kappa = 0.40$ Å^{-1} profiles are identical in all regions which indicates that the HE method correctly deals with electrostatic pressure contributions evaluated by Ewald summation.

Then, let us compare the pressure profiles obtained with the HE method and with different cutoffs of the electrostatic potential. The profiles are based on the 17 ns simulation where the Particle Mesh Ewald (PME) technique was used. Figure 4.5 shows the total pressure profiles ($P_L - P_N$) calculated using cutoffs of 8, 12, 16 and 20 Å as well as the Ewald-profile. Note that the pressure in the aqueous phase ($|z| > 26$ Å) is not isotropic in any of the profiles, which indicates that the simulation does not contain enough water for the aqueous phase to obtain bulk properties away from the bilayer. In Fig. 4.5 we see that the profile obtained with a cutoff of 8 Å deviates from the Ewald-profile by up to -400 bar in the aliphatic part of the bilayer ($|z| < 14$ Å), and in the aqueous phase ($|z| > 26$ Å) the deviations amount to about 100 bar. One expects that by increasing the cutoff, thereby including electrostatic interactions at larger distances, the cutoff-profiles will approach the Ewald-profile. This is indeed the case in all regions when the cutoff is increased from 8 Å to 12 Å, but surprisingly, Fig. 4.5 also shows that increasing the cutoff further to 16 Å and 20 Å causes an increase in the deviations from the Ewald-profile in some regions, while other regions only show small deviations. Thus, the electrostatic pressure profile does not

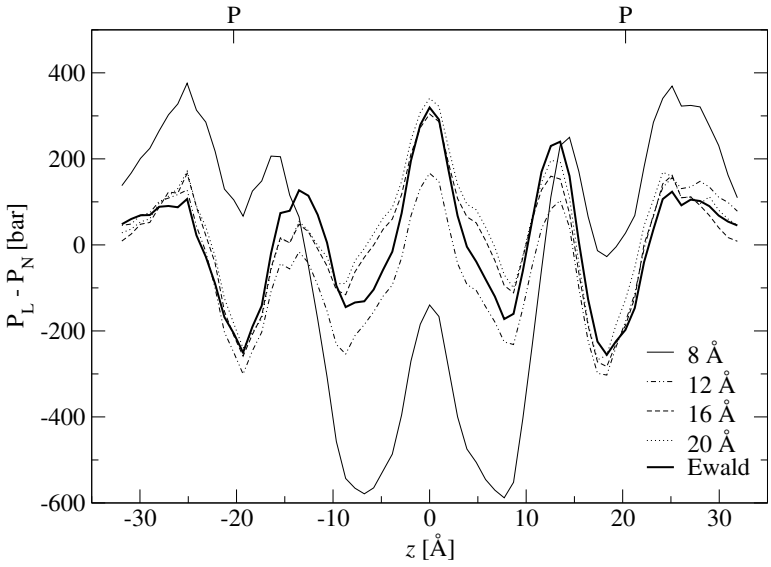
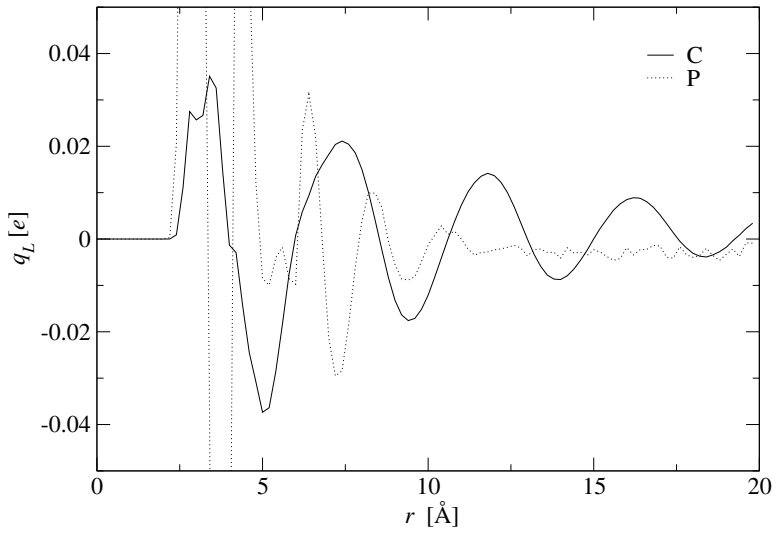


Figure 4.5: Total pressure profiles (4 point running average) calculated using cutoffs of 8, 12, 16 and 20 Å as well using HE method (Ewald). The maxima in the phosphorus density are marked with the letter P on the top axis of the graph. The standard error for all the pressure profiles are of the same magnitude as the standard error for the IK-profile in Fig. 4.3.

simply converge, in a monotonic fashion as the cutoff is increased. The pressure profiles in Fig. 4.5 result in surface tensions (γ) of 20±6 mN/m (cutoff 8 Å), 28±4 mN/m (cutoff 12 Å), -8±4 mN/m (cutoff 16 Å), -22±4 mN/m (cutoff 20 Å) and -6±2 mN/m (Ewald). The surface tension is calculated using $\gamma = \int_{-L^z/2}^{L^z/2}(\mathrm{P_N} - \mathrm{P_L})dz$ and the standard error for the average surface tension is estimated from the surface tensions in uncorrelated 250 ps time blocks. These values also reflect the non-monotonic convergence behavior in the system. A similar cutoff dependence is found for the energy of an ionic crystal, where it is well known that the electrostatic energy oscillates strongly as the cutoff is increased due to radial charge ordering. The analogy to ordered ionic structures is substantiated by the existence of a non-uniform lateral charge profile $q_L(r)$ in the bilayer (cf. Fig. 4.6(a)). The lateral charge profile at $r = r'$, calculated for any reference atom, gives the total charge in a cylindrical slab of height h, extending laterally from r' to $r' + \delta r$ away from the reference atom. Thus, oscillations in $q_L(r)$ indicate lateral (or 'in-slab') charge ordering in the system which suggests that the energy and pressure from 'in-slab' interactions will oscillate with increasing cutoff. In Fig. 4.6(a) the solid line shows the average lateral charge profile calculated for aliphatic carbon atoms with $|z| < 14$ Å. In the calculation of $q_L(r)$ the region $|z| < 14$ Å was divided into 14 slabs of width 2 Å and for all carbons in each slab, $q_L(r)$ was calculated with contributions from all other atoms in the slab. In each slab the $q_L(r)$'s around all carbon atoms



(a)

Figure 4.6: The solid line is the lateral radial charge profile measured from aliphatic carbon atoms ($|z| < 14$ Å). The dotted line is the lateral radial charge profile measured from phosphorus atoms in the region 24 Å$> |z| > 16$ Å. See text for details. As in the simulations, the electrostatic interactions between atoms that interact through covalent bonds or valence angle interactions were excluded in the calculation of the lateral charge profile.

were averaged to give one $q_L(r)$ for that slab and these $q_L(r)$'s were then averaged over the 14 slabs to give the solid line in Fig. 4.6(a). The figure shows that the bilayer possess lateral charge ordering which persists over the whole range considered. This explains why the cutoff-profiles converge toward the Ewald-profile in an oscillatory manner as seen in Fig. 4.5. Although 'in-slab' electrostatic interactions give an important lateral pressure contribution since the lateral components of both $\mathbf{f}_{ij}$ and $\mathbf{r}_{ij}$ are relative large and the whole pressure is assigned to one slab, the 'out-of-slab' electrostatic interactions may also have significant lateral components which are governed by $q_L(r)$ in other slabs. The dotted line in Fig. 4.6(a) shows $q_L(r)$ for phosphorus calculated in the region 24 Å$> |z| > 16$ Å using a slab width of 2 Å. Comparing $q_L(r)$ for aliphatic carbon atoms with $q_L(r)$ for phosphorus atoms (Fig. 4.6(a) solid and dotted lines, respectively) we see that $q_L(r)$ varies through the bilayer which is reflected in the quite complex cutoff dependence of the pressure profile. Adding to the complexity of the cutoff dependence of the pressure profile, one should keep in mind that the 'out-of-slab' contributions also contributes to $\mathrm{P_N}$. These may add or cancel contributions in $\mathrm{P_L}$ and therefore, the cutoff dependence of $\mathrm{P_L} - \mathrm{P_N}$ may be different from the one of $\mathrm{P_L}$.

Note that $q_L(r)$ for phosphorus converges towards a negative charge which means that the charge density in the phosphorus-rich region of the bilayer is negative. This indicates a non-uniform charge density in the direction normal to the bilayer giving rise to the electrostatic potential profile (see e.g. Refs. [19] and [143]).

We have found that increasing the electrostatic cutoff will not necessarily make the cutoff-profile approach the Ewald-profile. This behavior can be explained in terms of charge ordering in the system which, as for ordered ionic systems, means that the results are sensitive to the cutoff distance. By using e.g. charge-group based cutoff one can accommodate for some of the shortcomings of the brute force cutoff method. The charge-group based cutoff method would be straightforward to implement in pressure profile calculations both for the IK and H expressions, and would, in effect, damp the oscillations in the lateral charge profile. Therefore, pressure profiles calculated using the charge-group based cutoff are expected to be less cutoff dependent compared to the brute force cutoff profiles. A thorough investigation of different methods for treating electrostatic in MD simulations was carried out by Anézo et al. [4].

We would like to emphasize that although the cutoff-profiles in Fig. 4.5 deviate from the Ewald-profile, all the qualitative features of the Ewald-profile are conserved in the cutoff-profiles when using a reasonably long electrostatic potential cutoff (16 and 20 Å). However, since the simulation is conducted with Ewald summation, the cutoff-profiles are approximations to the Ewald-profile. Thus, even though we have found that pressure profiles in the DPPC lipid bilayer, calculated with a large cutoff from Ewald-simulations, are qualitatively correct, other systems may have a different charge structure and therefore may have a different cutoff dependence which is not known *a priori*.

In addition to allowing for calculating long range electrostatic contributions to the pressure profile, we would like to propose another interesting application of the H expression. Some MD simulation codes, such as GROMACS [18, 100], calculate the contribution to the bulk pressure from non-bonded (pair) interactions in a single loop over particles (see the GROMACS manual `www.gromacs.org/documentation/paper_manuals.php`). As described by Lindahl and Edholm [98], pressure profile calculations using the IK expression make the simulation (in GROMACS) several times slower since the IK expression has to be evaluated in the inner loop, where the positions of both particles in a pair are known. This requirement is, in essence, the reason why we cannot use the IK expression to calculate the reciprocal space contribution to the pressure profile. However, as we have seen above this is not an issue when using the H expression, since we only need the total virial contribution from particle i and the position of i in order to calculate the H pressure profile. Thus, the H expression should also allow

for calculating the pressure profile contributions from pair interactions in the outer loop (or a separate single loop) of an MD code. This should be an attractive alternative to the current pressure profile implementation in, e.g., GROMACS.

4.5 Conclusion

In this chapter we have addressed two central issues concerning pressure profile calculations, namely the arbitrariness in the choice of integration contour and the treatment of long-range electrostatic interactions.

Regarding the choice of integration contour we find that the Irving-Kirkwood (IK) and Harasima (H) expressions give qualitatively very similar pressure profiles for a DPPC lipid bilayer. The deviations between the IK and H profiles are below ± 50 bar in most regions. However, in systems with strong density oscillations, such as liquids close to a wall, it has been reported that the IK and H pressure profiles have regions where they deviate qualitatively [153]. Therefore, in such regions, the local pressure is not uniquely defined and a detailed interpretation of the local pressure must be encumbered with great uncertainty. Thus, with a given system, a comparison of the pressure profiles obtained by the IK and H expressions is advisable in order to check the uniqueness of the pressure profile.

Concerning the treatment of long-range electrostatic interactions, we have shown how the H expression for the local pressure can be used to include the electrostatic contributions to the lateral components of the local pressure tensor calculated from the Ewald sum through the so-called Harasima-Ewald-method (HE method). We used the HE method to calculate pressure profiles for a DPPC lipid bilayer and the pressure profile is found to be independent of the partitioning between the real space and the reciprocal space contributions to the Ewald sum.

The pressure profile can also be calculated using a finite cutoff for the electrostatic interactions. This approach is consistent if the simulations are also conducted with the same cutoff, but if the simulations are conducted using Ewald summation, such a pressure profile is inconsistent. The effect of this inconsistency was investigated by comparing pressure profiles calculated with different electrostatic potential cutoffs (8, 12, 16 and 20 Å) with the Ewald-profile. We found that the cutoff-profiles approach the Ewald-profile in a non-monotonic fashion which was attributed to the existence of long-range charge ordering in the bilayer. Even though the variations in the pressure profile with cutoff are found to be rather subtle, the qualitative features of the cutoff-profiles are similar to the Ewald-profile when a relatively large cutoff is used (16 and 20 Å). Further, considering the uncertainty in the pressure profile caused by the arbitrariness in the choice of integration contour, the deviations between the cutoff-profiles and the Ewald-profile are

not serious as long as a relatively large cutoff is used. However, other systems may have a different charge structure compared to a DPPC bilayer and therefore the cutoff dependence of the pressure profile may be different.

For systems that are simulated using Ewald summation, the HE method can be used for pressure profile calculation without investigating the cutoff dependence of the pressure profile, regardless of the charge structure in the system. Thus, calculating the pressure profile by the HE method has great advantages compared to calculations using an electrostatic potential cutoff.

5 Structural transitions in the ABC transporter BtuCD

Based on: J. Sonne, C. Kandt, G.H. Peters, F.Y. Hansen, M.Ø. Jensen and D.P. Tieleman. Simulated nucleotide binding and release in the ABC transporter BtuCD: Coupled motions of the nucleotide binding domains and transmembrane domains. In preparation.

5.1 Introduction

ABC transporters are membrane spanning proteins containing the ATP Binding Cassette (ABC) [69]. They use energy from adonesine triphosphate (ATP) binding and hydrolysis to drive active transport of a wide variety of compounds across biological membranes. There are 48 different transporters in humans and 80 in the gram-negative bacterium *E. coli*, transporting different compounds such as anions, lipids and amino acids [20, 40, 71]. Malfunction of ABC transporters can have dramatic pathological consequences [44, 126] and account for multidrug resistance [22], e.g., human tumor cell resistance to cytotoxic drugs used in chemotherapy [59, 72]. Thus, it is of great interest to understand in detail how ABC transporters use ATP to facilitate transport.

ABC-transporters have two membrane spanning domains (MSDs) and two ABC domains, which are often referred to as nucleotide binding domains (NBDs). The ATP hydrolysis takes place in an NBD dimer sandwich where highly conserved sequences from both NBD monomers contribute to nucleotide binding and hydrolysis. These sequences are the Walker A motif, Walker B motif, the H loop, Q loop and the LSGG motif [40, 86]. The latter sequence is so specific to ABC transporters that it is referred to as the ABC signature motif [71, 103]. Four crystal structures of *E. coli* MalK, the NBD part of the maltose transporter, provide insight into the rearrangements induced in the NBDs of ABC transporters by binding and hydrolysis of ATP [36, 104].

In all ABC transporters, such rearrangements are expected to propagate to the MSDs thereby facilitating transport. The basis of this propagation is not well understood [102] since the only transporter structure that has

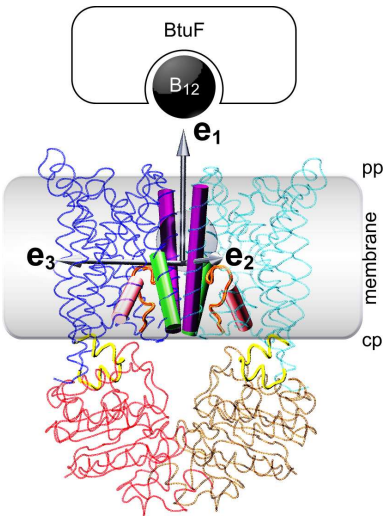


Figure 5.1: Crystal structure of the *E. coli* vitamin B_{12} transporter BtuCD. Side view of the BtuCD crystal structure [103] with a schematic illustration of the membrane and the periplasmic binding protein BtuF with bound vitamin B_{12}. The cytoplasmic and periplasmic side of the membrane are indicated by "cp" and "pp", respectively. The two transmembrane domains (MSDs) are rendered in a tube representation and colored in blue and cyan, respectively. The nucleotide binding domains (NBDs) are also rendered in a tube representation and are colored in red and brown, respectively. The L-helix motifs of the MSDs that make important non-covalent contacts to the NBDs are rendered as yellow tubes. Several other structural elements are highlighted for later reference: The C-terminal helices of the MSDs are shown as green cylinders. The fifth transmembrane helices of both MSDs (TM5s) are shown as violet cylinders. The eighth transmembrane helices of both MSDs (TM8s) are shown as pink cylinders. The orange tubes show the extended pore-lining motifs connecting the TM2 and TM3 helices (pore loops). The putative translocation path of vitamin B_{12} goes in between the two transmembrane domains along the $\mathbf{e}_1$-axis into the central cavity (transparent gray sphere). The cytoplasmic gate is thought to be comprised of the cytoplasmic ends of the two TM5 helices [101, 103]. The three gray arrows denoted $\mathbf{e}_1$,$\mathbf{e}_2$ and $\mathbf{e}_3$ show the principal components of the inertia tensor of the two MSDs. $\mathbf{e}_1$ will also be referred to as the pore axis. The illustration was prepared using VMD [73].

been solved in both an open and a closed state is the lipid transporter MsbA [34, 35]. In both MsbA structures, however, the fold and orientation of the NBDs suggest that these structures are subject to crystallization artifacts [40, 86, 149]. Another ABC transporter for which the full structure (MSDs and NBDs) is available is the *E. coli* vitamin B_{12} importer BtuCD [PDB 1L7V, [103]] shown in Fig. 5.1. The BtuCD crystal structure is determined at a 3.20 Å resolution and has cyclovanadate bound in both catalytic sites.

The BtuCD structure inspired the development of the so-called BtuCD-based transport model [101, 103], which is illustrated in Fig. 5.2(b). This model assumes that the BtuCD crystal structure is close to the physiological resting state in which the cytoplasmic gate is closed. In this resting state

the periplasmic gate is open such that the substrate can enter the pore upon release from the binding pocket of the periplasmic binding protein BtuF (see Fig. 5.1). ATP binding brings the NBDs into a tight dimeric configuration, and through a toggle switch motion, the cytoplasmic gate is opened and the substrate is translocated. The cytoplasmic gate closes when the hydrolysis products are released and transporter returns to its resting state.

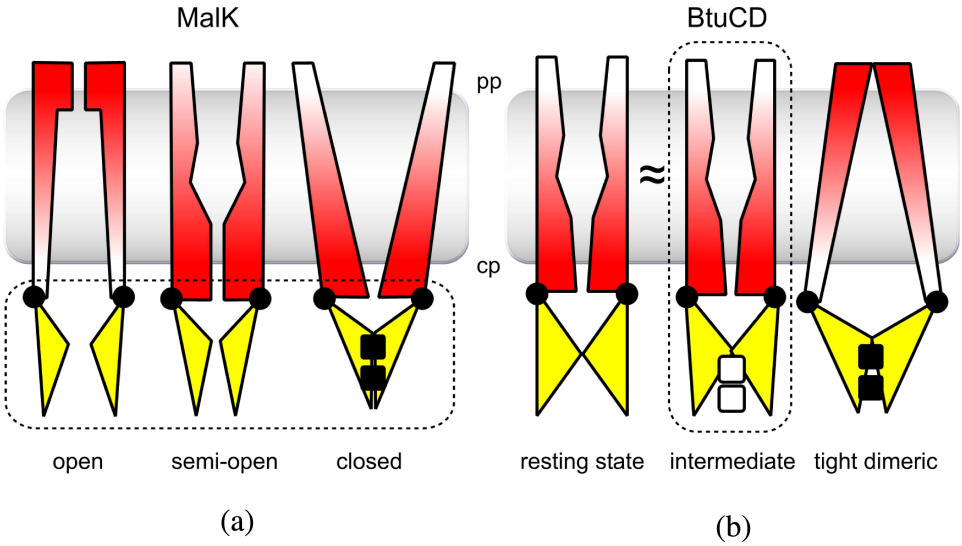


Figure 5.2: Two proposed transport models in ABC transporters. (a) The MalK-based model: ATP binding closes the cytoplasmic gate [36]. (b) The BtuCD-based model: ATP binding opens the cytoplasmic gate through a toggle switch motion [103]; The MSDs are colored in red and white. The red ends of the MSDs correspond to the closed end of the transporter pore, while the white ends correspond to the open end of the transporter pore. The NBDs are yellow. The solid black squares symbolize ATP, the solid white squares symbolize cyclo-vanadate and the solid gray squares in the background symbolize the membrane. The dashed boxes mark structures that are known from X-ray crystallography and therefore structures outside these boxes are hypothetical. We have illustrated the MalK-based model differently from previous reports [36, 40] to stress that the similar fold of the MalK and the NBDs of BtuCD suggests that ATP binding occurs in equivalent regions [40, 86].

The so-called MalK-based model, illustrated schematically in Fig. 5.2(a), is founded on crystal structures of two ATP-free and one ATP-bound conformations of *E. coli* MalK [36]. This model predicts that ATP binding induces the NBDs to go into the closed MalK configuration (PDB 1Q12, resolution 2.60 Å), but in contrast to the BtuCD-based model, this closes the cytoplasmic gate while opening the periplasmic gate. Opening of the periplasmic gate allows the substrate to enter the transporter. Subsequent release of hydrolysis products translocates the substrate and resets the transporter into the resting state (open MalK configuration, PDB 1Q1E, resolution 2.90 Å). The crystal structure of semi-open MalK (PDB 1Q1B, resolution 2.80 Å) is thought to correspond to a reaction intermediate in the ATP hydrolysis. Ir-

respective of which transport model applies to BtuCD, it is unclear whether the structural rearrangements in the MSDs are large enough for the transport substrate to leave the transporter by diffusion or if peristaltic forces, generated by the MSDs on the substrate, are necessary for transport [103].

The transport mechanism in ABC transporters has previously been investigated by molecular dynamics (MD) simulations. Since structures of full ABC transporters are scarce, some studies simulated the NBD part of different transporters and focused on rearrangements in these NBD dimers upon ATP binding and hydrolysis to ADP [27, 28, 85]. Oloo and Tieleman [120] simulated the full BtuCD transporter in its native membrane environment with and without ATP and found that ATP binding draws the NBDs closer to each other. At the same time the TM5 helices moved closer together at the periplasmic side, but no significant opening of the cytoplasmic gate occurred. The purpose of this computational study is to characterize the nucleotide induced rearrangements in the MSDs of BtuCD to better understand the mechanism that drives substrate transport. In conventional MD simulations, the atomic positions are propagated with a time step of a few femtoseconds, which limits the total simulation time to 10-100 nanoseconds for transmembrane protein simulations [5]. This timescale is often not long enough to probe biologically relevant processes, such as the domain rearrangements we are looking for in BtuCD. To allow for a more exhaustive exploration of global protein conformations in MD simulations, essential dynamics sampling techniques have been developed [42]. Another method for gaining insight into the long time scale conformational dynamics of proteins in computer simulations is the elastic network model, which has proved successful in reproducing experimental data despite its simplicity [6, 9, 11, 47, 63]. Within the elastic network model, Zheng and Brooks [160] showed that structural transitions between two states of a protein, for example induced by ligand binding, can be modeled in a perturbational variant of the elastic network model.

To simulate conformational transitions in proteins, both the essential dynamics and perturbed elastic network methods require some knowledge of the target structure. This is inherently problematic in BtuCD since only one crystal structure has been solved [103]. In order to circumvent this problem, we assume that the three *E. coli* MalK crystal structures each corresponds to a physiological state also occurring in the physiological function of *E. coli* BtuCD, and in turn derive from these structures two directions that take MalK from the semi-open to closed and semi-open to open state. We then use these directions as a bias in our MD simulations of BtuCD. Similarly, the structures of MalK are used as guidance for the perturbations in the BtuCD elastic network. Using these approaches, the nucleotide induced transitions in BtuCD are simulated devoid of sampling problems, however, we cannot interpret the rearrangements found in the NBDs since these are closely coupled to the simulation input. Therefore, our analysis is focused

on the response of the MSDs quantified in terms of changes in the distance between the L-helix motifs of each MSD domain, changes in the MSD-MSD tilt angle and structural rearrangements in the cytoplasmic gate region.

Our results indicate that transition to the loose dimeric state, which we think is equivalent to nucleotide release, increases the distance between the L-helix motifs, decreases the tilt angle between the two MSDs and opens the cytoplasmic gate in BtuCD. Transition to the tight dimeric state has the opposite effect. Comparing with the mechanisms schematically illustrated in Fig. 5.2, the MalK-based transport model appears to be valid also for BtuCD and the idea of a toggle switch motion in the MSD-NBD transmission interface, as suggested in the BtuCD-based model, cannot be confirmed. The structural rearrangements in the cytoplasmic gate region, induced by simulated nucleotide release are, however, not large enough to allow vitamin B_{12} to diffuse into the cytoplasm. Therefore we speculate that peristaltic forces, generated by the MSDs, could play an important role when B_{12} is excluded from the central cavity to the cytoplasm.

5.2 Simulation details

5.2.1 Elastic network calculations

Equilibrium vibrations in MalK and BtuCD were modeled in the anisotropic network model (see Sec. 2.2.1). In our elastic network calculations we used a cutoff distance of 13 Å and a spring constant of 1 kcal/(mol Å^2) as suggested in previous ANM studies [6].

The transient influence of ATP was investigated using a perturbational variant of the elastic network [160] (see Sec. 2.2.2). In this approach the effect of ligand binding is accounted for by introducing a low number of C_α-C_α distance constraints as a perturbation to the network energy. This perturbation gives a response vector that describes how all C_αs in the protein respond to changing a few C_α-C_α pair distances. Zheng and Brooks [160] tested the method on 22 proteins for which two crystal structures, that each represent one conformational state, are known. They chose ten C_α pairs for the perturbation that cross the network cutoff and have the most significant distance change between the two structures. This approach is not viable for BtuCD since only one crystal structure is available. Therefore, we find the relevant C_α pairs from crystal structures of MalK. From the alignment table presented by Davidson and Chen [40] we map the perturbation sites from MalK to the equivalent sites in the NBDs of BtuCD. The perturbations in the NBDs of BtuCD then mimic nucleotide binding and release events and the response in the MSDs can in turn be monitored. The perturbation sites in MalK were confined to regions conserved among the NBDs in ABC transporters and therefore the same is true after mapping the sites onto BtuCD. In MalK, analysis of the C_α-C_α pair distance

changes between the semi-open and the open form suggests that this transition should be modeled by introducing perturbations between the signature motif and the Walker A motif in the opposing NBD. The transition from the semi-open to the closed form of MalK, require perturbations between the Q-loops in the two NBDs, the signature motif and the Walker A motif in opposing NBDs as well as between the Q-loop and the signature motif in the same NBD. By introducing these perturbations, we tested and confirmed that the predictions of the structural transitions in MalK significantly improved compared to the unperturbed elastic network which is the success criterion proposed by Zheng and Brooks [160]: Going from the semi-open to the open form of MalK, the dot product between the normalized response vector and the normalized (experimental) structural transition vector is 0.7. For the transition from the semi-open to the closed form of MalK, corresponding dot product is 0.6. For both transitions the biggest overlap with any eigenvector from standard elastic network and the structural change is 0.4 (eigenvector 1) which is much lower than the overlaps of the perturbed network. Therefore both the opening and closing transitions in MalK are successfully modeled by the perturbed network when the perturbations are introduced in functionally relevant regions only.

Confining the selection of C_α pairs to the functionally relevant regions did not reduce the performance compared to selecting the pairs from the whole protein. We have implemented the perturbed network method in the elastic network code currently available for download at the website of the Jernigan group (`http://ribosome.bb.iastate.edu/software.html`).

5.2.2 Molecular dynamics simulations

For the molecular dynamics (MD) simulations, the crystal structure of *E. coli* BtuCD (PDB 1L7V) was placed in a palmitoyloleonylphosphatidylethanolamine (16:0/18:1c9, POPE) lipid bilayer consisting of 326 united atom lipids and solvated with 25 000 simple point charge [SPC [66]] water molecules as described by Kandt et al. [90]. 20 chloride ions were added for electroneutrality. The full system contained $\sim$ 100 000 atoms and mimics the plasma membrane of *E. coli* [144]. All residues were protonated according to the standard protonation scheme in GROMACS [18, 100]. The MD simulations were performed in GROMACS using the GMX3.2.1 ffgmx force field for the protein. Constant pressure was maintained by a Berendsen barostat with a time-constant of 4 ps, a compressibility of $4.5 \cdot 10^{-5}$ bar^{-1} and a target pressure of 1 bar. The pressure coupling in the direction parallel to the bilayer normal was carried out independently from the two other directions. A Berendsen thermostat, with a time-constant of 0.1 ps, assured an average temperature of 310 K. The temperature bath was independently coupled to the protein, the bilayer and the water plus chloride ions. The LINCS algorithm [67] was used to constrain all bond lengths except water where the

SETTLE [113] algorithm was used. Constraining the bonds allowed us to use an integration timestep of 2 fs. Electrostatic interactions were evaluated using three dimensional Particle Mesh Ewald summation [39, 50] and van der Waals interactions were evaluated directly from the 12-6 Lennard-Jones potential using a cutoff of 1.4 nm. The Lennard-Jones potential was modified by a switching function in the whole interaction range. The BtuCD system was equilibrated for 8 ns after which the C_α Root Mean Square Deviation (RMSD) of BtuCD converged at 0.23 nm relative to the crystal structure. Equilibrium data was collected during the succeeding 4 ns.

5.2.3 Essential dynamics sampling

We used linear acceptance essential dynamics sampling [41, 42] to explore the directions that are relevant for nucleotide binding and release more efficiently. The input is a direction in which the algorithm should increase the sampling. For this input direction we calculated the transition vectors from the semi-open to the open and from the semi-open to the closed structures of MalK and used the (x, y, z)-elements for the C_αs close to functionally important and highly conserved regions, i.e., residues 30-44, 77-83, 124-138, 153-161 and 188-194 in both NBDs of BtuCD. The efficiency of the essential dynamics sampling technique allowed us to run ten simulations biasing in the direction corresponding to nucleotide binding and ten simulations biasing in the direction corresponding to nucleotide release. The essential dynamics simulations were started from the same system configuration extracted after 12 ns of equilibration of the system described in Sec.5.2.2. Velocities were reassigned from a Maxwell-Boltzmann distribution with a temperature of 310 K in each simulation. The transitions to the tight dimeric structure took $\sim$75 ps, while the transitions to the loose dimeric structure took $\sim$150 ps. These simulation times are artificial since the MD simulations are biased to speed up sampling.

Both the perturbed network and the essential dynamics simulations only give the direction in which the protein deforms upon the simulated nucleotide binding and release, but not the amplitude of the motion. Thus, we define the tight/loose dimeric structure of BtuCD as having the lowest RMSD from homologue parts of the closed/open MalK structure. The RMSD between the equilibrated BtuCD structure and homologue parts of the closed and open MalK crystal structures are 4.8 and 9.0 Å, respectively. In the biased MD simulations, the minimal RMSD between BtuCD and homologue parts of the closed MalK crystal structure, defining the tightly dimerized state of BtuCD, is $\sim$3.8 Å. Correspondingly, the minimal RMSD between BtuCD and homologue parts of the open MalK crystal structure, defining the loose dimeric state, is $\sim$4.5 Å.

5.3 Results

5.3.1 Overall response of BtuCD

The idea of using MalK as a template is illustrated in Fig. 5.3 and the main results of this study are contained in Fig. 5.4, which summarizes the overall response in BtuCD. In Fig. 5.3 the start and end structures of BtuCD taken from one representative MD simulation for the loose and tight NBD dimerization transitions. The corresponding MalK crystal structures, used to model these transitions, are also shown. The structural overlay illustrates that, on a global scale, essential dynamics sampling in MD simulations can be used to mimic the effect of nucleotide binding and release. Figure 5.4(a)

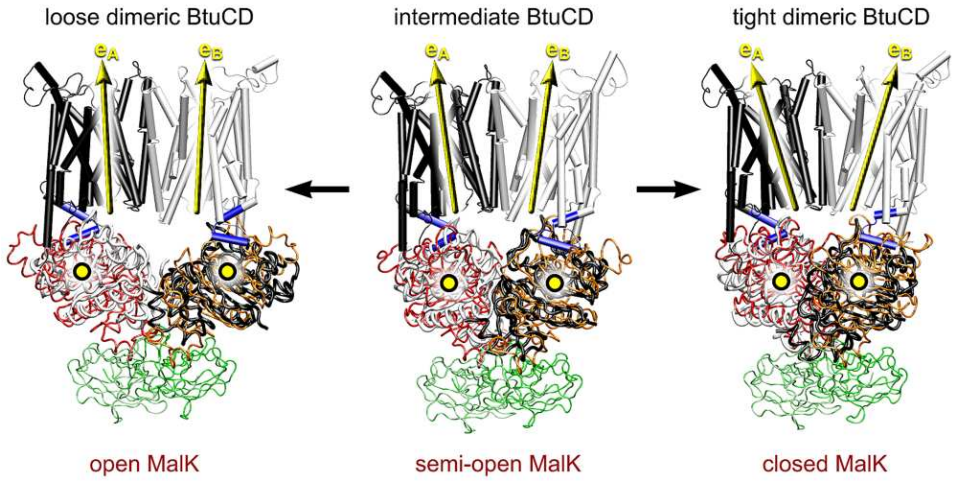


Figure 5.3: Simulated nucleotide binding and release in BtuCD as predicted by MD simulation with essential dynamics sampling. Visualization of the overall structural rearrangements in BtuCD for the transition from the "intermediate BtuCD" to the "loose dimeric BtuCD" structure and the transition from the "intermediate BtuCD" to the "tight dimeric BtuCD" structure. The "intermediate BtuCD" structure is the X-ray structure equilibrated for 12 ns in a POPE bilayer with water on both sides of the bilayer (see Sec. 5.2.2). The BtuCD is rendered in black and white cartoon and tubes except for the L-helix motifs which are blue. The tight and loose dimeric structures of BtuCD are defined as the trajectory frames having the lowest root mean square deviation from homologue parts of the corresponding closed and open MalK structures (see Sec. 5.2.1). The BtuCD structures are overlayed with the corresponding MalK crystal structures (orange, red and green tubes). The regulatory domains of MalK (green) are not present in the BtuCD structure. The filled yellow circles schematically illustrate the centers of mass of each NBD in BtuCD and the yellow arrows marked $\mathbf{e}_A$ and $\mathbf{e}_B$ are inertia (long) axes for each MSD. The illustration was prepared using VMD [73].

shows the distance between the two NBDs in BtuCD during the transition to the loose and tight dimeric state. As expected, both the elastic network and the MD simulations predict that the transition to the loose dimeric state of BtuCD increases the NBD-NBD distance and that this distance is

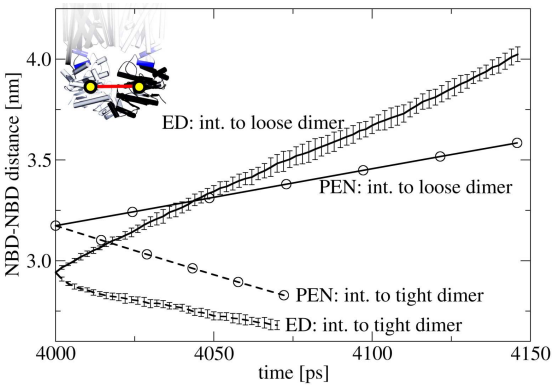
decreased upon transition to the tight dimeric state. The perturbed elastic network calculations start from the BtuCD crystal structure while the biased MD simulations start from a BtuCD structure equilibrated for 12 ns. This accounts for the mutually different initial values in the perturbed elastic network calculations and MD simulations.

The rearrangements in the NBDs are thought to propagate into the MSDs via the L-helices located in the NBD-MSD interface [101, 103]. The inter L-helix distances for the two transitions in BtuCD are shown in Fig. 5.4(b). Comparing with Fig. 5.4(a), the distance between the L-helices in opposing MSDs is seen to be strongly coupled to the NBD-NBD distance. Both the perturbed network and the MD simulations predict that a transition to the loose dimeric state, corresponding to nucleotide release, increases the L-helix distance. Nucleotide binding has the opposite effect. Figure 5.4 (c) shows the response in the MSD-MSD tilt upon nucleotide binding and release. The MSD-MSD tilt, defined as the angle between the MSD inertia (long) axes shown in Fig. 5.3, increases upon tight NBD dimerization (nucleotide binding) and decreases upon transition to the loose dimeric state (nucleotide release).

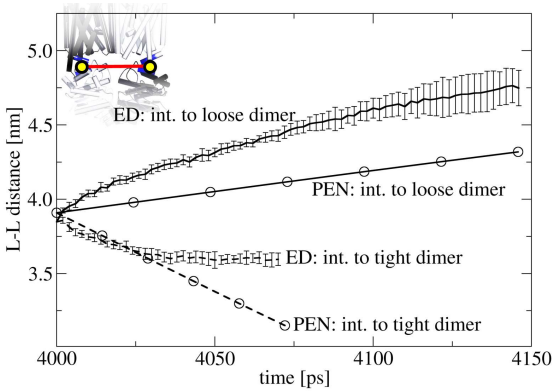
5.3.2 Response in the cytoplasmic gate region

To monitor the response in the cytoplasmic gate region we will first characterize the putative translocation path (pore) through the transmembrane part of BtuCD in the crystal structure. We find two possible cytoplasmic exit paths and in Fig. 5.5(a), the pore radii for these are shown as a function of x_1, i.e. the position projected onto the pore axis ($\mathbf{e}_1$ in Fig. 5.1). In Fig. 5.5(b) we have visualized the two pores corresponding to the plots in panel (a) and the surrounding BtuCD structure. The gray pore, referred to as the TM5 exit, corresponds to the dotted curve in Fig. 5.5(a). It essentially lies parallel to $\mathbf{e}_1$ sandwiched between transmembrane helices 5 (TM5) of the two MSD monomers. The other exit path, referred to as the C-terminal exit, departs from the TM5 exit path at $x_1 = -0.6$ nm and exits into the cytoplasm sandwiched in between the pore loop, the TM5, TM8, and the C-terminal helices. The rotational symmetry of the MSD dimer allows for a second C-terminal exit at the other side of the TM5 exit path, providing two equivalent C-terminal exits in BtuCD. In the following, all descriptions of the response in the C-terminal exit region refer to both these exits.

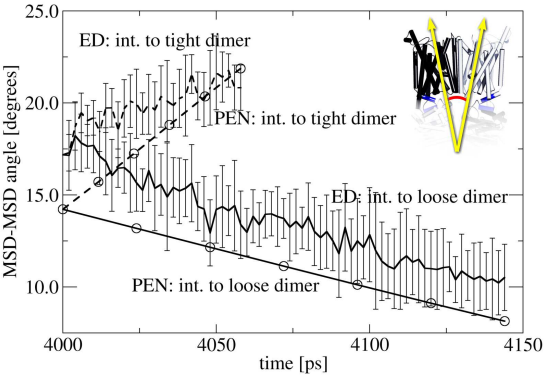
The structural response of BtuCD around the two possible cytoplasmic exits as predicted by the perturbed elastic network is shown in Fig. 5.6. Panels (a-b) and (c-d) illustrate the transition from the crystal structure to the loose and tight dimeric structures of BtuCD, respectively. For illustrational purposes we have extrapolated the response beyond the structure that has the lowest root mean square deviation (RMSD) from the corresponding MalK structure. Thus, the dramatic deformation of the L-helix motifs



(a)



(b)



(c)

Figure 5.4: Coupling between NBD motions and MSD response. Structural responses associated with nucleotide binding and release in BtuCD. (a) The distance between the two NBD centers of mass. (b) The distance between the centers of mass of the two L-helix motifs. (c) The tilt between the two MSDs defined as the angle between the inertia axes of each MSD ($\mathbf{e}_A$ and $\mathbf{e}_B$ in Fig. 5.3). The insets illustrate the measured quantities in the protein structure. In all panels the solid black curves correspond to the transition to the loose dimeric structure, while the transition to the tight dimeric structure are shown in dashed lines. The circles mark results of the perturbed elastic network (PEN) which, to ease the comparison, were adjusted to the same timescale as the MD simulations. The error bars on the MD results are the standard deviations obtained from ten simulations carried out for each of the two transitions. There are no error bars on the perturbed elastic network results, since the method is deterministic and only one starting structure was used (the crystal structure). The perturbed elastic network calculations start from the BtuCD crystal structure (int.) while the biased MD simulations start from an equilibrated BtuCD structure (int.). The abscissa axes have the unit of time which is somewhat artificial since the MD simulations are biased to speed up sampling and the notion of time is absent in the perturbed elastic network calculations.

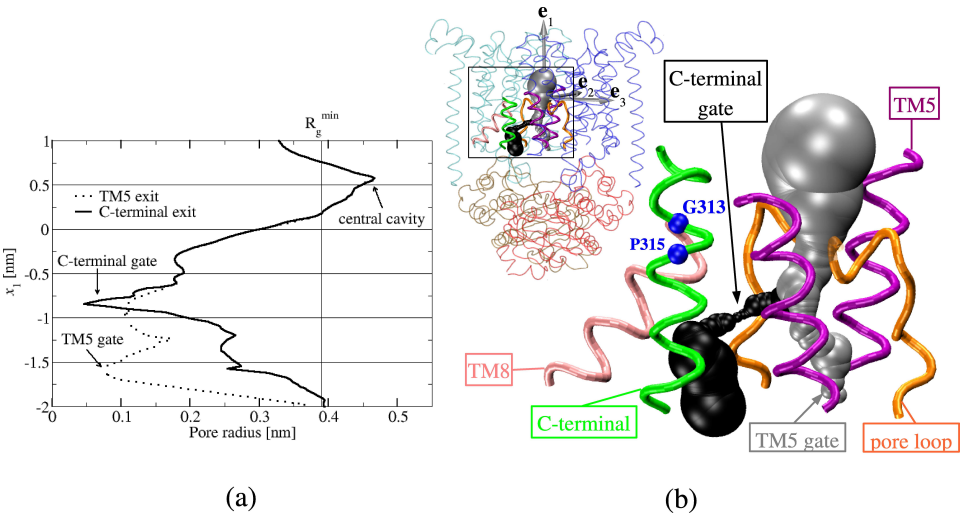


Figure 5.5: Two possible cytoplasmic exits in BtuCD. (a) Pore radii as a function of the position projected on the pore axis $\mathbf{e}_1$. The graph shows the radii for two possible pores through the crystal structure of BtuCD: The TM5 exit path (dotted) and the C-terminal exit path (solid). We have indicated the dimensions of B_{12} by the vertical line at R_g^{min} which is the minimal radius of gyration of B_{12}. The pore radii are calculated using the program Hole [136]. (b) Visualization of the two pores and key structural elements of the surrounding protein: The gray surface is the TM5 exit path, the black surface is the C-terminal exit path. The protein is rendered as follows: TM5 helix (purple tube), TM8 (pink tube), pore loop (orange tube) and C-terminal helix (green tube). We represent the molecular hinge residues G313 and P315 [147] as blue spheres. The inset between panels (a) and (b) shows the two pores in the full BtuCD structure. The gray arrows show the principal components of the inertia tensor of the MSDs. $\mathbf{e}_1$ is the pore axis referred to on the ordinate axis in panel (a). The rectangle shows the part of the protein magnified in panel (b). Panels (a) and (b) are approximately aligned. The visualizations were prepared using VMD [73].

in panels (c-d) is merely a consequence of this exaggeration and the end-structures in Fig. 5.4, that fulfill the minimal RMSD criterion (see Sec. 5.2) do not show such unphysical behavior. Fig. 5.6 shows that the TM5 helices (purple cylinders) change their tilt relative to the pore axis upon nucleotide binding and release. The change in the tilt primarily takes place in the $\mathbf{e}_3$-direction, but not the $\mathbf{e}_2$-direction. The transition to the loose dimeric structure causes the TM8 helices (pink cylinder in Fig. 5.6) and the pore loop (orange tubes) to retract from the pore center line. Transition to the tight dimeric structure has the opposite effect. The response of the C-terminal helix upon transition to the loose dimeric configuration does not affect the size of the C-terminal exit (see Fig. 5.6). However, upon transition to the tight dimeric configuration of BtuCD, the C-terminal helix moves into the C-terminal exit path, thus contracting the pore.

Rupture of the MSD-NBD interface in the MD simulations prevents us from exaggerating these responses and consequently the structural re-

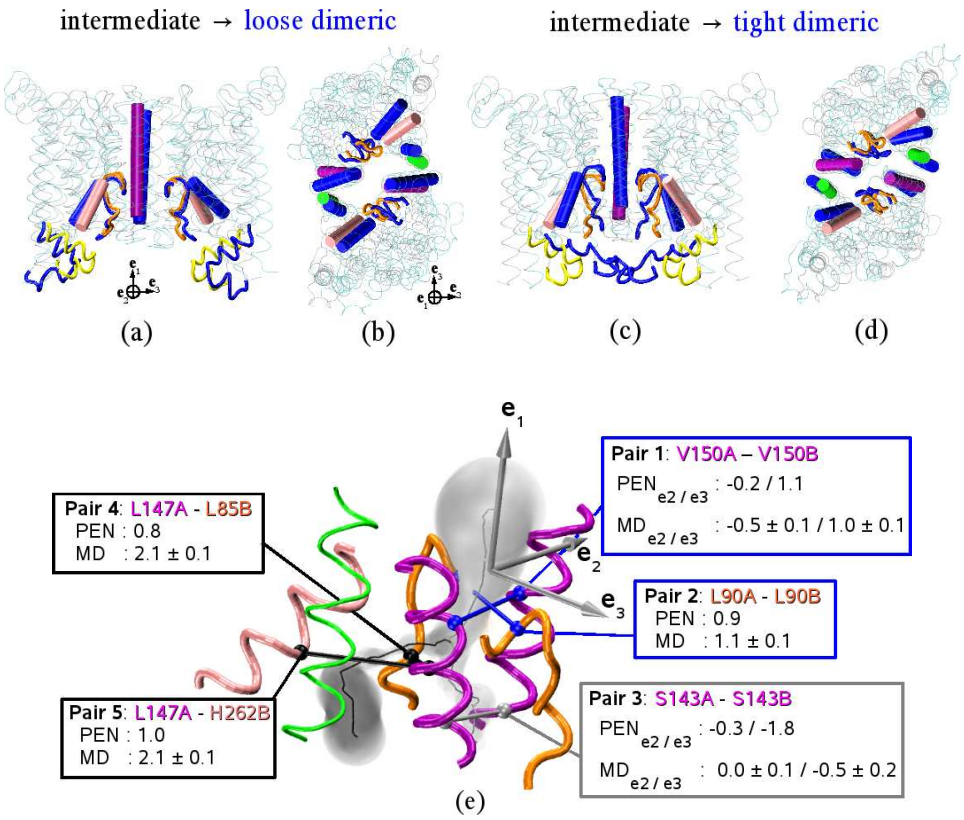


Figure 5.6: Nucleotide induced structural response in the cytoplasmic gate region. Illustration of the transitions from the BtuCD crystal structure (intermediate) to the loose and tight dimeric states as predicted by the perturbed elastic network in side view and cytoplasmic view. The transitions are shown as an overlay of the BtuCD crystal structure with the loose and tight response structures. Transition to the loose dimeric state: (a) Side view; (b) Cytoplasmic view. Transition to the tight dimeric state: (c) Side view; (d) Cytoplasmic view. The key structural elements in the cytoplasmic exit regions of the intermediate BtuCD structure are rendered as purple cylinder (TM5), green cylinder (C-terminal helix), pink cylinder (TM8) and orange tube (pore loop). The L-helix motifs are rendered as yellow tubes. The corresponding structural elements in the loose and tight dimeric structures are all rendered in blue. To increase clarity, not all structural elements are shown in both the side view and cytoplasmic view panels. To show the key structural elements more clearly, the MSD crystal structure is rendered in semi-transparent gray tubes, while the loose and tight dimeric structures are rendered in semi-transparent cyan tubes. For clarity the NBDs are not shown. Note that the amplitude of the responses are exaggerated as to illustrate the motions of the different structural elements more clearly. (e) Distance changes for five C_α-C_α pairs for the transition from the tight to the loose dimeric configurations as predicted by the perturbed elastic network (PEN) and the biased MD simulations (MD). For the MD simulations, we have indicated estimates of the standard deviations of the mean distance changes. For this analysis, the tight and loose dimeric structures of BtuCD are defined as the trajectory frames having the lowest root mean square deviation from homologue parts of the corresponding closed and open MalK structures (see Sec. 5.2.1). Thus, we do not use the exaggerated structures shown in (a-d). The protein structure essentially has the same orientation as in Fig. 5.5(b) as also illustrated by the three principal components of the inertia tensor for the MSDs (gray arrows). Illustrations were prepared using VMD [73].

arrangements are too subtle to show visually.

In principle, we could have quantified the changes in the pore dimensions using Hole as in Fig. 5.5. However, this approach would not make it possible to compare the predictions of the MD simulations with the perturbed elastic network calculations, since the latter technique only contains C_α atoms. Thus, we analyzed the structural rearrangements numerically by monitoring changes in C_α-C_α distances in the two cytoplasmic gate regions. For this analysis we choose C_α atoms belonging to pore-lining residues in the bottleneck regions below the central cavity (cf. Fig. 5.5). Further, we choose the C_α atoms so that the line connecting the C_α atoms in a pair approximately crosses the pore center line in the crystal structure. With these choices we believe it is reasonable to use the C_α pair distances as indicators of the pore dimensions and of the directions of the responses in key regions of the pore. The distance analysis for the MD simulations and the perturbed network is presented in Fig. 5.6(e) for the transition from the tight dimeric state to the loose dimeric state, thus excluding the intermediate structure used for the modeling. For pairs 1 and 3, we present the $\mathbf{e}_2$ and $\mathbf{e}_3$ components of the distance change, which allows for a more fruitful rendering of the response than would be provided by monitoring the net distance changes. Pairs 1 and 2 are located just below the central cavity, but before the C-terminal exit path splits from the TM5 exit path. The distance changes for pair 1 indicate that upon nucleotide release, the two TM5 helices approach each other in the $\mathbf{e}_2$ direction and that their separation increase in the $\mathbf{e}_3$ direction. The distance changes in the $\mathbf{e}_3$ direction are larger than in the $\mathbf{e}_2$ direction. The results obtained for pair 2 indicate that the two pore loops retract from the pore center line upon transition to the loose dimeric state. Pair 3, measures the TM5-TM5 helix distance in the TM5 gate region. For this pair, the response in the $\mathbf{e}_3$ direction clearly dominates the response in the $\mathbf{e}_2$ direction and interestingly the MD simulations do in fact predict that there is no significant response to the nucleotide (un)binding events in the latter direction. Pairs 4 and 5, located in the C-terminal exit region, clearly indicate that transition to the loose dimeric state increases the distance between the TM5 helix and the pore loop and between the TM5 helix and the TM8 helix. Note that the above interpretations of the distance changes remain unchanged even when the confidence is increased to, e.g., a 95 % level. The distance analysis for the perturbed elastic network confirms the overall picture of the response presented in Fig. 5.6(a-d), which indicates that the pairs chosen for the analysis describe the response consistently with visual inspection of the perturbed elastic network trajectories. Further, since the distance analysis from perturbed elastic network calculations and the MD simulations agree qualitatively and the changes in the pairs distances are significant in the MD simulations, Fig. 5.6(a-d) should also illustrate the average effect of the transitions as predicted by the MD simulations.

Thus, in summary, the predictions of MD simulations and the perturbed

elastic network agree qualitatively. The response of the TM5 exit region can best be described as a tilt motion of the TM5 helices in the $\pm\mathbf{e}_3$ directions. Therefore, it appears that nucleotide (un)binding does not facilitate opening of the TM5 exit. In contrast, the transition from the tight to the loose dimeric state of BtuCD appears to be coupled to opening of the C-terminal exit. However, the largest distance increase in the C-terminal exit is merely $\sim$2 Å.

5.4 Discussion

It is commonly accepted that tight dimerization of the NBDs in ABC transporters is induced by ATP binding, whereas the transition to the loose dimeric state may be coupled to release of hydrolysis products [36, 40, 103]. These transitions are thought to cause structural rearrangements in the MSDs that facilitate transport, however, this coupling is not well understood [102]. In this chapter, we study such global structural rearrangements in the *E. coli* vitamin B_{12} importer BtuCD by computer simulations. Due to the long time scales involved, global rearrangements are difficult to capture in conventional MD simulations. Therefore, instead of including the nucleotides explicitly in MD simulations, as previously done by Oloo and Tieleman [120], we modeled the effects of nucleotide binding and release in a perturbed elastic network model and in biased MD simulations. We used experimentally observed transitions in MalK as templates for perturbing and biasing the NBDs in BtuCD and then monitored the response of the MSDs.

5.4.1 Which transport model?

When the NBDs of BtuCD dissociate into the loose dimeric configuration, the distance between the L-helix motifs, connecting the two MSDs to the NBDs, increases and the tilt angle between the two MSDs decreases (cf. Fig. 5.4). The transition to the tight dimeric state has the opposite effect. Comparison with Fig. 5.2 indicates that these results are consistent with the MalK-based transport model [36, 40]. Hence, the BtuCD-based model and the concept of a toggle-switch mechanism cannot be confirmed. The results of the perturbed elastic network and the biased MD simulations are qualitatively similar, which shows some universality in our prediction and establishes the usefulness of the computationally efficient perturbed network approach. However, compared to the elastic network, the atomistic MD force field includes important anharmonic effects [47] and accounts for the properties of the individual residues. Modeling the transition in MD simulations could therefore give information on specific interactions across the transmission interface and allow for testing mutations. We focused on

comparing the MD results with the perturbed elastic network and with proposed transport mechanisms in the literature.

Recently Vergani et al. [155] reported that the tightly dimerized NBD conformation of human epithelial cystic fibrosis transmembrane conductance regulator (CFTR) channels corresponds to the CFTR-channel open state. In ABC exporters and importers, this state is thought to be equivalent to an MSD configuration where the periplasmic gate is open while the cytoplasmic gate is closed [155]. Using site directed spin labeling and electron paramagnetic resonance spectroscopy on the *E. coli* lipid exporter MsbA Dong et al. [46] found that ATP binding closes the cytoplasmic gate. If there is a common transport mechanism among ABC importers and exporters as suggested by van der Does and Tampe [149], these experimental results support the MalK-based transport model and hence also our results. It should be noted, however, that such a common mechanism among ABC importers and exporters has not been confirmed yet [20]. In the BtuCD MD study by Oloo and Tieleman [120], the periplasmic side of the translocation path closes when introducing explicit ATP in MD simulations, while little happens in the cytoplasmic gate region. The latter result is consistent with our observations, while the closing of the periplasmic side is not. The discrepancy could reflect a sampling problem in the work of Oloo and Tieleman, since limitations in computer time allowed the authors to carry out only one ATP binding simulation. Alternatively, the discrepancy could indicate a problem in our approach.

Since our study confirms the MalK-based transport model, a valid concern is to what extend the transitions we find in BtuCD are biased by the MalK-like input. To this end it is important to note that MalK and the NBDs of BtuCD have almost complete sequence identity in the regions that are involved in ATP binding and hydrolysis, i.e., the Walker A, Walker B, signature motifs as well as the H and Q loops. Further, the similar folds of the BtuCD NBDs and MalK assure that the architectures of the binding sites in the two proteins are almost identical. Thus, the consensus in the literature is that the structural changes in the above mentioned regions, induced by nucleotide binding and release, are the same in the two transporters [40, 86, 103]. If the maltose transporter and BtuCD facilitate transport in different ways, it must therefore be a consequence of differences in the overall protein structure since the local events in the binding sites are the same. To minimize the MalK-like biasing we therefore confined the perturbation sites and the elements of the displacement vector to functionally relevant regions, i.e., the Walker A, Walker B, signature motifs as well as the H and Q loops [40]. This ensures that the perturbations and the biasing the MD simulations in BtuCD are localized to regions involved in the physicochemical events that account for the structural changes in MalK. Since the same physicochemical events take place in BtuCD, the bias introduced by the MalK-like input is minimal.

5.4.2 Cytoplasmic gate opening.

Our analyses of the putative transport pore in Sec. 5.3.2 suggest that there may exist two types of cytoplasmic exit paths, which we refer to as the TM5 exit and the C-terminal exit, respectively (cf. Fig. 5.5). If nucleotide binding or release were responsible for opening the TM5 exit, one would expect the response in the cytoplasmic ends of the TM5 helices to be directed in the $\mathbf{e}_2$-direction (cf. Fig. 5.1). Such a motion is, however, observed in neither the perturbed network nor the biased MD simulations, which is consistent with the MD simulations of Oloo and Tieleman [120]. The structural elements that make up the C-terminal exits, on the other hand, are found to retract from the pore center line upon transition to the loose dimeric state of BtuCD in both the perturbed elastic network model and in the biased MD simulations. Therefore, release of hydrolysis products, which could reset the transporter to the resting state, seems to be responsible for opening the C-terminal exit. However, the observed gate opening is not large enough to accommodate vitamin B_{12}. Thus, it is interesting that the C-terminal helix, which lines the C-terminal exit, contains a molecular hinge, i.e., a GXP motif [147] [see Fig. 5.5(b)]. This hinge (G313-P315), located just below the central cavity at $x_1 \sim -0.2$ nm, could provide considerable structural flexibility in the C-terminal exit region. Indeed, the crystallographic B-factors do increase down through the C-terminal helix, but it is not possible to attribute this increase to the GXP motifs or to the function of BtuCD without further investigations.

5.4.3 Two transport scenarios

Compared to the schematic transport models presented in the literature [36, 40, 103, 149], the magnitude of the cytoplasmic gate opening we find is strikingly small. Moreover, the structural changes in the C-terminal exit are not large enough to allow B_{12} to diffuse from the central cavity and out into the cytoplasm. This finding points to one of the following two transport scenarios:

(I) The structural response from nucleotide binding and release are in fact as small as predicted here. In this case, the modest opening indicates that peristaltic forces from the MSDs pushing B_{12} towards the cytoplasm could be important. The changes in the MSD tilt upon nucleotide release drive the periplasmic ends of the MSDs, which could generate a force pushing B_{12} towards the cytoplasm via the C-terminal exit. In the lines of these speculations, the C-terminal helix could act as a "passive hatch", which moves to the side as B_{12} propagates through the C-terminal. The term "passive hatch" refers to the fact that such a motion of the C-terminal helix is not seen in the present study and is therefore expected to be independent of nucleotide binding and release. If this is true, and the C-terminal helix is

simply pushed to the side by B_{12}, it is important that the C-terminal helix is moved easily, which could suggest a possible hinge role of the GXP-motifs in the C-terminal helices.

(II) The structural rearrangements we find are too small and larger rearrangements, coupled to events that are not accounted for in the present study, control transport. For example, the structural responses to nucleotide binding and release could be altered by the presence of the periplasmic binding protein BtuF. The presence of binding proteins stimulate ATPase activity [40] and in the schematic transport mechanism presented by Locher [101], BtuF sends a "signal" from the periplasmic side of BtuCD to the hydrolysis sites which triggers ATP hydrolysis. The signaling between BtuF and the hydrolysis sites could involve changes in the transporter structure that in turn could change the response to nucleotide binding and release. This hypothesis could be tested by the same techniques as employed in the present study once the structure of the BtuF-BtuCD complex is solved experimentally or modeled.

The findings of the present study are consistent with scenario (I) but we cannot rule out scenario (II) without further investigations.

5.5 Conclusion

We have studied the response in the BtuCD structure upon nucleotide binding and release with computer simulations techniques. We find that the responses to nucleotide binding and release in BtuCD are consistent with the schematics of the MalK-based transport model, which therefore could apply for BtuCD. Hence, the BtuCD-based model and the notion of a toggle-switch motion cannot be confirmed from our simulations. The observed coupling between the nucleotide binding domains and the membrane spanning domains may be relevant for all ABC transporters, because of the conservation of nucleotide binding domains and the shared role of ATP in ABC transporters. The structural rearrangements we find in the cytoplasmic exit region, however, do not provide enough space for B_{12} to diffuse from the transporter pore out into the cytoplasm. We therefore speculate that peristaltic forces, generated by changes in the tilt of the membrane spanning domains, are needed to exclude B_{12} from the central cavity of the transporter pore. In future work it would be interesting to study the influence of the periplasmic binding protein BtuF on the nucleotide induced rearrangements in BtuCD. Further, to better understand how the structural rearrangements in the NBDs propagate across the NBD-MSD interface, it would be interesting to carry out biased MD simulations, as described in the present study, where residues in the transmission interface have been subject to mutations.

6 Conclusion

This chapter is, in part, based on abstracts for manuscripts A, B and C (see Manuscripts & publications page 75).

In this thesis, computer simulation techniques were used to study three different aspects of the nature of cell membranes. Since these three aspects follow the sectioning of the thesis in chapters, I have found it natural to structure the conclusion in the same manner and at the end discuss some more general aspects of simulating biosystems.

In **Chapter 3** we demonstrate that molecular dynamics (MD) simulations of dipalmitoylphosphatidylcholine (DPPC) lipid bilayers using the commonly employed CHARMM27 force field reproduces fluid (L_α) phase bilayer properties only when a positive surface tension is applied. The tension required to obtain fluid phosphatidylcholine bilayers is not universal and has to be adjusted based on experimental data such as the area per lipid. Therefore, bilayer simulations in the tensionless isothermal-isobaric (NPT) ensemble are appealing since they do not rely on experimental input data, provided that the force field is optimized for simulations in this ensemble. Unfortunately, NPT simulations using the CHARMM27 force field give highly ordered, gel-like bilayer structures at temperatures and pressures, where the bilayers should be in the biologically relevant fluid phase [53, 82]. After 15 ns the area per lipid for DPPC bilayers is in fact close to that of the DPPC gel phase [82], i.e., $\sim$48 Å^2. We argue, that the bilayer ordering cannot solely be attributed to finite size effects and is therefore largely a consequence of shortcomings in the force field for NPT bilayer simulations. To obtain fluid phase properties of DPPC lipid bilayers represented by the CHARMM energy function in the NPT ensemble, we reparameterized the atomic partial charges in the lipid head group and the first three acyl chain methylene groups. The new charges were determined from a quantum calculation of the electron structure using both the Mulliken method and the restricted electrostatic potential (RESP) method, where partial charges are determined by a fit to the electrostatic potential as calculated from the electron distribution. The new partial charges were tested in MD simulations

of fully hydrated DPPC lipid bilayers together with otherwise unchanged CHARMM27 force field parameters in the NPT ensemble. The replacement of the CHARMM27 partial charges with the new Mulliken charges had no significant effect on the area per lipid. In contrast, simulations with the new RESP charges show significant improvements compared with simulations using the original CHARMM27 force field, manifested in a resulting area per lipid of 60.4 ± 0.1 Å^2, which conforms closely to the experimental value of 64 Å^2 [116]. Also, the simulated order parameter profile and electron density profile are in better agreement with experiments than simulations using the original CHARMM27 force field. The results also indicate that there is room for additional optimizations of out new parameters and that such optimizations should focus on the glycerol backbone and upper acyl chain regions. In any case, our new CHARM27 based parameters make it possible, for the first time, to simulate the biologically more interesting fluid phase of DPPC bilayers in all-atom simulations in the NPT ensemble.

In **Chapter 4** we address two central and unresolved problems in pressure profile calculations. The first problem is that the pressure profile is not uniquely defined since the expression for the local pressure involves an arbitrary choice of an integration contour. We investigated two different integration contours leading to the Irving-Kirkwood (IK) and Harasima (H) expressions for the local pressure tensor. For these choices we find that the pressure profile is almost independent of the contour used, which indicates that the local pressure is well-defined for a fluid DPPC bilayer represented by the new parameter set found in chapter 3. For the comparison it was necessary to truncate the electrostatic interactions energies and describe them as a sum of pair interactions, which is required in the IK method. This requirement naturally introduces the second problem we address, namely how to include electrostatic interactions in pressure profile calculations when the simulations are conducted using the Ewald summation technique for the electrostatic interactions. Based on the H expression, we present a method for calculating the contribution to the lateral components of the local pressure tensor from electrostatic interactions evaluated by the Ewald summation technique. Using this new expression, we show that pressure profiles calculated with an electrostatic potential truncation (cutoff) from simulations conducted with Ewald summation depend on the cutoff in a subtle manner which is attributed to long-range charge ordering in the system. We also find that the pressure profiles calculated with relatively long cutoffs are qualitatively similar to the Ewald-profile for the DPPC bilayer studied here. The properties of the H expression that allow us to calculate the local pressure contribution from the Ewald sum should also allow for single loop pressure profile calculations. This should represent an attractive alternative to the current pressure profile implementation in, e.g., GROMACS [18, 100], which is based on the IK expression and slows down the simulations consid-

erably [98].

In **Chapter 5** we simulated the transmembrane protein BtuCD in a palmitoyloleonylphosphatidylethanolamine (POPE) lipid bilayer thought for mimicking the plasma membrane of *Escherichia coli* (*E. coli*). BtuCD is an adonesine triphosphate (ATP) binding cassette (ABC) transporter that uses ATP to drive active transport of vitamin B_{12} into the cytoplasm of *E. coli* [103]. The (dys)function of ABC transporters accounts for, e.g., cystic fibrosis and multi-drug resistance, for example tumor cell resistance to anticancer drugs [40, 44, 126]. Unfortunately, the mechanism facilitating substrate translocation in ABC transporters, which is of both fundamental and medicinal interest, is largely unknown. To study the nucleotide induced rearrangements in BtuCD, the effect of the nucleotides on the structure on the binding sites were modeled using perturbed elastic network calculations and biased MD simulations. These two techniques assure better sampling of the nucleotide induced configurational transitions in BtuCD than provided by conventional MD simulations. Both the perturbed elastic network calculations and the biased MD simulations predict that nucleotide release decreases the tilt between the two membrane spanning domains and appears to open the cytoplasmic gate. Nucleotide binding has the opposite effect. These findings are consistent with the MalK-based transport model, which therefore also appears to apply to BtuCD. The observed rearrangements in the cytoplasmic gate region do, however, not provide enough space for B_{12} to diffuse passively from the transporter pore into the cytoplasm, which could suggest that peristaltic forces, generated by changes in the tilt between the two membrane spanning domains, are needed to release B_{12} from the transporter pore. Owing to the conservation of nucleotide binding domains and the shared role of ATP in ABC transporters, the observed coupling between the nucleotide (un)binding events and the rearrangements in the membrane spanning domains may be relevant for all ABC transporters. The results of the perturbed elastic network calculations and the biased MD simulations are qualitatively similar, which shows some universality in our prediction and establishes the usefulness of the computationally efficient perturbed network approach.

In **MD simulations of biological systems**, there are three major challenges, namely to continously improve the force fields, to sample for longer times and to increase the system sizes. The enormous span of time/length scales in biological systems introduces a tradeoff between simulating in full atomistic detail and sampling long time/length scales. Thus, describing mesoscopic and macroscopic phenomena in a simulation often requires that the atomistic system description is abandoned with the consequence that some details are lost. This tradeoff, as well as the empirical nature of the force field, implies that there is no such thing as a *perfect* MD simulation and

therefore one may ask what is a *good* MD simulation? Except for obvious issues such as correct system setup, a *good* simulation should be designed to sample the time/length scales that are relevant for the phenomenon one wishes to study. Therefore, the design of the simulation includes choosing the appropriate level of detail in the model, developing or finding a corresponding force field as well as choosing the simulation type. Thus, any simulation inevitably ignores some (unimportant) aspects of the studied phenomenon (on either short or long time/length scales) and therefore there will always be properties that are represented reasonably well in a simulation, while others are not. This implies that not all quantities that can be calculated from a simulation have a sound interpretation and therefore one should be careful that the analyses of a simulation are consistent with the choice of model, the simulation type and the sampled time/length scales.

Since the discussion in the preceding paragraph may be somewhat abstract, I will use the results from chapter 5 to illustrate some of the points stated above. Recognizing that the timescale of nucleotide induced rearrangements in BtuCD are on the limit of those accessible to MD simulations, we mimic the effect of nucleotide binding and release instead of introducing the nucleotides explicitely. In terms of the tradeoff between time/length scales and simulation detail, this approach prioritizes the former. The disadvantage is that the nucleotides are "man made" and therefore the rearrangements in nucleotide binding sites are controlled by the simulation input, not by explicit atomic interactions. Therefore, in the quest for better sampling of the global rearrangements in BtuCD, we sacrifice the detailed information on the rearrangements in the nucleotide binding sites that simulations with explicit nucleotides would provide. Obviously, this implies that interpreting the rearrangements in the nucleotide binding sites would be inappropriate. Further, the methods employed to mimic the effect of the nucleotides confound the time scales in the simulations and therefore estimating kinetic data directly from the simulations times would not yield sound results. The advantage of our approach is mainly that it allows us to carry our multiple simulations of the (un)binding events. Thus, we can carry out statistical analysis on the simulations and avoid interpreting events observed only once. For this discussion, it is interesting that our predictions, based on multiple simulations with modeled nucleotides, on some points contradict the predictions of Oloo and Tieleman [120], which are based on a single MD simulation where ATP was introduced explicitely. At the present time it is not possible to determine, which results are more correct, that is, whether the contradictions are caused by artifacts in our modeled nucleotides or by insufficient sampling in the simulation of Oloo and Tieleman [120]. In either case, this example illustrates that it is not trivial to choose between two simulation types for a given phenomenon.

Manuscripts & publications

A. J. Sonne, F.Y. Hansen, and G.H. Peters. Methodological problems in pressure profile calculations for lipid bilayers. Chem. Phys., **122** 124903 (2005)

B. J. Sonne, M.Ø. Jensen, F.Y. Hansen, L. Hemmingsen and G.H. Peters. A parameter study of dipalmitoylphosphatidylcholine lipids. Submitted.

C. J. Sonne, C. Kandt, G.H. Peters, F.Y. Hansen, M.Ø. Jensen and D.P. Tieleman. Simulated nucleotide binding and release in the ABC transporter BtuCD: Coupled motions of the nucleotide binding domains and transmembrane domains. In preparation.

D. E. Oloo, J. Sonne and D.P. Tieleman. Insights into the mechanism of coupling between the ATP-binding cassette and the transmembrane domains of BtuCD, the vitamin B_{12} ABC transporter. In preparation.

Dansk resumé

Molekyldynamiske (MD) computersimuleringer benyttes i dag i stor udstrækning til at studere forskellige typer systemer på submikroskopiske længdeskalaer. I denne afhandling med titlen *Computer simulations of lipid bilayers and proteins* beskrives MD simuleringer af biologiske membraner og proteiner. I en MD simulering løses Newtons bevægelsesligning numerisk for en samling af op til nogle hundrede tusinde atomer. Hovedresultatet af en sådan simulering er tidsudviklingen af atompositioner og en analyse heraf, giver mulighed for at studere bl.a. protein-dynamik og -funktion i atomar detalje. Arbejdet kan naturligt opdeles i tre hovedemner:

Modelparametre: MD simuleringer er baseret på et sæt empirisk bestemte modelparametre, som beskriver atomare vekselvirkninger. Kvaliteten af disse modelparametre er bestemmende for, i hvor stor udstrækning simuleringerne repræsenterer virkeligheden. Den membrantype vi har studeret var for kompakt i MD simuleringer sammenlignet med eksperimentelle resultater og derfor er modelparametrene for fosfolipider, som udgør hovedbestanddelen af cellemembraner, blevet optimeret. Denne optimering har muliggjort simuleringer af membraner med realistiske fysiske egenskaber, hvilket er en vigtig forudsætning for videre studier af komplekse biologiske membraner.

Trykprofiler i membraner: En membran er $\sim$5 nm tyk, og på denne korte afstand varierer trykket med flere hundrede bar. Disse variationer i trykket betegnes som membranens trykprofil. Der er fremsat teorier om, at nogle anæstesifremkaldende stoffer virker ved at ændre trykprofilen i cellemembranen. Denne hypotese er svær at verificere, da trykprofiler ikke kan bestemmes eksperimentelt. MD simuleringer er derfor et vigtigt redskab til at undersøge trykprofiler og deres relation til anæstesi. I den sammenhæng er et analyseprogram, der kan beregne trykprofiler fra MD simuleringer, blevet udviklet. Programmet benytter en ny beregningsmetode, der tillader en mere præcis bestemmelse af trykprofilen i forhold til tidligere benyttede metoder.

Transportmekanismen i ABC transporteren BtuCD: ABC transportere, så som BtuCD, udnytter energi fra adenosintrifosfat (ATP) til at transportere en række molekyler ind og ud af levende celler. Størstedelen af cystisk fibrose patienter bærer defekte ABC transportere, som derfor ikke er i stand til at varetage deres transportfunktion. ABC transportere er desuden bestemmende for celleresistens imod antibiotika og imod celledræbende lægemidler som benyttes i kemoterapi. Det er derfor af stor fundamental og medicinsk interesse at forstå detaljerne i funktionen af ABC transportere. BtuCD transporterer B_{12}-vitamin ind gennem plasma membranen på *E. coli* bakterien og er en af de eneste ABC transportere, for hvilken den rumlige struktur er kendt. Vores computersimuleringer har bidraget til en forståelse af transportcyklusen i BtuCD, som kan være relevant for ABC transportere generelt, grundet deres strukturelle og funktionelle ligheder.

Bibliography

[1] Alejandre, J., D. Tildesley, and G. Chapela. 1995. Molecular dynamics simulation of the orthobaric densities and surface tension of water. *J. Chem. Phys.* 102:4574–83.

[2] Allen, M. P., and D. J. Tildesley. 1989. Computer Simulations of Liquids. Clarendon Press, Oxford.

[3] Andresen, T. L., S. S. Jensen, and K. Jørgensen. 2005. Advanced strategies in liposomal cancer therapy: Problems and prospects of active and tumor specific drug release. *Prog. Lipid Res.* 44:68–97.

[4] Anézo, C., A. H. de Vries, H. Holtje, D. P. Tieleman, and S. Marrink. 2003. Methodological issues in lipid bilayer simulations. *J. Phys. Chem. B.* 107:9424–33.

[5] Ash, W. L., M. R. Zlomislic, E. O. Oloo, and D. P. Tieleman. 2004. Computer simulations of membrane proteins. *Biochim. Biophys. Acta Biomembr.* 1666:158–189.

[6] Atilgan, A. R., S. R. Durell, R. L. Jernigan, M. C. Demirel, O. Keskin, and I. Bahar. 2001. Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys. J.* 80:505–15.

[7] Bahar, I., A. R. Atilgan, M. C. Demirel, and B. Erman. 1998. Vibrational dynamics of folded proteins: significance of slow and fast motions in relation to function and stability. *Phys. Rev. Lett.* 80:2733–6.

[8] Bahar, I., B. Erman, R. L. Jernigan, A. R. Atilgan, and D. G. Covell. 1999. Collective Motions in HIV-1 Reverse Transcriptase: Examination of Flexibility and Enzyme Function. *J. Mol. Biol.* 285:1023–1037.

[9] Bahar, I., and R. L. Jernigan. 1997. Inter-residue Potentials in Globular Proteins and the Dominance of Highly Specific Hydrophilic Interactions at Close Separation. *J. Mol. Biol.* 266:195–214.

[10] Bahar, I., and R. L. Jernigan. 1998. Vibrational Dynamics of Transfer RNAs: Comparison of the Free and Synthetase-bound Forms. *J. Mol. Biol.* 281:871–884.

[11] Bahar, I., A. Wallqvist, D. G. Covell, and R. L. Jernigan. 1998. Correlation between Native-State Hydrogen Exchange and Cooperative

Residue Fluctuations from a Simple Model. *Biochemistry.* 37:1067–1075.

[12] Bandyopadhyay, S., J. C. Shelley, and M. L. Klein. 2001. Molecular dynamics study of the effect of surfactant on a biomembrane. *J. Phys. Chem. B.* 105:5979–5986.

[13] Baoukina, S. V., and S. I. Mukhin. 2004. Bilayer membrane in confined geometry: Interlayer slide and entropic repulsion. *J. Exp. Theor. Phys.* 99:875.

[14] Bayly, C. I., P. Cieplak, W. D. Cornell, and P. A. Kollman. 1993. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.* 97:10269–10280.

[15] Benz, R. W., F. Castro-Roman, D. J. Tobias, and S. H. White. 2005. Experimental validation of molecular dynamics simulations of lipid bilayers: a new approach. *Biophys. J.* 88:805–817.

[16] Berendsen, H. J. C., and S. Hayward. 2000. Collective protein dynamics in relation to function. *Curr. Opin. Struc. Biol.* 10:165–169.

[17] Berendsen, H. J. C., J. P. M. Postma, W. F. van Gunsteren, and J. Hermans. 1981. *In* Intermolecular forces. B. Pullman, editor. Reidel, Dordrecht, Holland. 331.

[18] Berendsen, H. J. C., D. van der Spoel, and R. van Drunen. 1995. GROMACS: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* 91:43–56.

[19] Berger, O., O. Edholm, and F. Jahnig. 1997. Molecular dynamics simulations of a fluid bilayer of dipalmitoylphosphatidylcholine at full hydration, constant pressure, and constant temperature. *Biophys. J.* 72:2002–13.

[20] Biemans-Oldehinkel, E., M. K. Doeven, and B. Poolman. 2006. ABC transporter architecture and regulatory roles of accessory domains. *FEBS Lett.* 580:1023–1035.

[21] Boal, D. 2002. Mechanics of the Cell. Cambridge University Press.

[22] Borst, P., and R. O. Elferink. 2002. Mamalian ABC transporters in health and disease. *Annu. Rev. Biochem.* 71:537–592.

[23] Bretscher, M. S. 1985. The Molecules of the Cell Membrane. *Sci. Am.* 253:100–108.

[24] Brooks, B. R., R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. 1983. CHARMM: a program for macromolecular energy, minimisation, and dynamics calculations. *J. Comput. Chem.* 4:187–217.

[25] Brown, D., and S. Neyertz. 1995. A general pressure tensor calculation for molecular dynamics simulations. *Mol. Phys.* 84:577–95.

[26] Callisen, T. H., and Y. Talmon. 1998. Direct imaging by Cryo-TEM shows membrane break-up by phospholipase A2 enzymatic activity. *Biochemistry.* 37:10987–10993.

[27] Campbell, J. D., S. S. Deol, F. M. Ashcroft, I. D. Kerr, and M. S. P. Sansom. 2004. Nucleotide-dependent conformational changes in HisP: molecular dynamics simulations of an ABC transporter nucleotide-binding domain. *Biophys. J.* 87:3703–3715.

[28] Campbell, J. D., and M. S. P. Sansom. 2005. Nucleotide binding to the homodimeric MJ0796 protein: A computational study of a prokaryotic ABC transporter NBD dimer. *FEBS Lett.* 579:4193–4199.

[29] Cantor, R. S. 1997. The lateral pressure profile in membranes: A physical mechanism of general anesthesia. *Biochemistry.* 36:2339–44.

[30] Cantor, R. S. 1998. The lateral pressure profile in membranes: a physical mechanism of general anesthesia. *Toxicol. Lett.* 100:451.

[31] Cantor, R. S. 2002. Membranes - size distribution of barrel-stave aggregates of membrane peptides: Influence of the bilayer lateral pressure profile. *Biophys. J.* 82:2520–2525.

[32] Cantor, R. S. 2003. Receptor desensitization by neurotransmitters in membranes: Are neurotransmitters the endogenous anesthetics? *Biochemistry.* 42:11891–11897.

[33] Carrillo-Tripp, M., and S. E. Feller. 2005. Evidence for a mechanism by which ω-3 polyunsaturated lipids may affect membrane protein function. *Biochemistry.* 44:10164–10169.

[34] Chang, G. 2003. Structure of MsbA from Vibrio cholera: A multidrug resistance ABC transporter homolog in a closed conformation. *J. Mol. Biol.* 330:419–430.

[35] Chang, G., and C. B. Roth. 2001. Structure of MsbA from E.coli: A homolog of the multidrug resistance ATP binding cassette (ABC) transporters. *Science.* 293:1793–1800.

[36] Chen, J., G. Lu, J. Lin, A. L. Davidson, and F. A. Quiocho. 2003. A Tweezers-like Motion of the ATP-Binding Cassette Dimer in an ABC Transport Cycle. *Mol. Cell.* 12:651–662.

[37] Chiu, S., M. Clark, V. Balaji, S. Subramaniam, H. L. Scott, and E. Jakobsson. 1995. Incorporation of surface tension into molecular dynamics simulation of an interface: a fluid phase lipid bilayer membrane. *Biophys. J.* 69:1230–1245.

[38] Colombo, G., S. J. Marrink, and A. E. Mark. 2003. Simulation of MscL gating in a bilayer under stress. *Biophys. J.* 84:2331–2337.

[39] Darden, T., D. York, and L. Pedersen. 1993. Particle mesh Ewald: an N.log(N) method for Ewald sums in large systems. *J. Chem. Phys.* 98:10089–92.

[40] Davidson, A. L., and J. Chen. 2004. ATP-binding Cassette Transporters in Bacteria. *Annu. Rev. Biochem.* 73:241–68.

[41] de Groot, B. L., A. Amadei, R. M. Scheek, N. A. J. van Nuland, and H. J. C. Berendsen. 1996. An extended sampling of the configurational space of HPr from E.coli. *Proteins.* 26:314–322.

[42] de Groot, B. L., A. Amadei, D. M. F. van Aalten, and H. J. C. Berendsen. 1996. Toward an exhaustive sampling of the configurational spaces of the two forms of the peptide hormone guanylin. *J. Biomol. Struct. Dyn.* 13:741–51.

[43] de Vries, A. H., A. E. Mark, and S. J. Marrink. 2004. Molecular dynamics simulation of the spontaneous formation of a small DPPC vesicle in water in atomistic detail. *J. Am. Chem. Soc.* 126:4488–4489.

[44] Dean, M., A. Rzhetsky, and R. Allikmets. 2001. The Human ATP-Binding Cassette (ABC) Transporter Superfamily. *Genome Res.* 11:1156–1166.

[45] DeLano, W. L. 2002. DeLano Scientific, San Carlos, CA, USA.

[46] Dong, J., G. Yang, and H. S. Mchaourab. 2005. Structural basis of energy transduction in the transport cycle of MsbA. *Science.* 308:1023–1028.

[47] Doruker, P., A. R. Atilgan, and I. Bahar. 2000. Dynamics of Proteins Predicted by Molecular Dynamics Simulations and Analytical Approaches: Application to a-Amylase Inhibitor. *Proteins.* 40:512–524.

[48] Douliez, J., A. Leonard, and E. J. Dufourc. 1995. Restatement of order parameters in biomembranes: calculation of C-C bond order parameters from C-D quadrupolar splittings. *Biophys. J.* 68:1727–1739.

[49] Egberts, E., S. Marrink, and H. J. C. Berendsen. 1993. Molecular dynamics simulation of a phospholipid membrane. *Eur. Biophys. J.* 22:423–436.

[50] Essmann, U., L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen. 1995. A smooth particle mesh Ewald method. *J. Chem. Phys.* 103:8577–93.

[51] Feller, S. E., and A. D. MacKerell Jr. 2000. An improved empirical potential energy function for molecular simulations of phospholipids. *J. Phys. Chem. B.* 104:7510–7515.

[52] Feller, S. E., and R. W. Pastor. 1996. On simulating lipid bilayers with an applied surface tension: periodic boundary conditions and undulations. *Biophys. J.* 71:1350–1355.

[53] Feller, S. E., and R. W. Pastor. 1999. Constant surface tension simulations of lipid bilayers: The sensitivity of surface areas and compressibilities. *J. Chem. Phys.* 111:1281–7.

[54] Feller, S. E., R. M. Venable, and R. W. Pastor. 1997. Computer simulation of a dppc phospholipid bilayer: Structural changes as a function of molecular surface area. *Langmuir.* 13:6555–6561.

[55] Feller, S. E., Y. Zhang, R. W. Pastor, and B. R. Brooks. 1995. Constant pressure molecular dynamics simulation: the Langevin piston method. *J. Chem. Phys.* 103:4613–21.

[56] Foloppe, N., and A. D. MacKerell, Jr. 2000. All-atom empirical force field for nucleic acids. I. parameter optimization based on small molecule and condensed phase macromolecular target data. *J. Comput. Chem.* 21:86–104.

[57] Frisch, M. J., G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. J. A. Montgomery, R. E. Stratmann, J. C. Burant, S. Dapprich, A. D. D. J. M. Millam, K. N. Kudin, M. C. Strain, O. Farkas, V. B. J. Tomasi, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K. Morokuma, D. K. Malick, A. D. Rabuck, J. B. F. K. Raghavachari, J. Cioslowski, J. V. Ortiz, B. B. S. A. G. Baboul, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, C. Gonzalez, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, J. L. Andres, C. Gonzalez, M. Head-Gordon, E. S. Replogle, , and J. A. Pople. 1998. Gaussian 98, Revision A.7. Gaussian, Inc., Pittsburgh PA.

[58] Goetz, R., and R. Lipowsky. 1998. Computer simulations of bilayer membranes: Self-assembly and interfacial tension. *J. Chem. Phys.* 108:7397–409.

[59] Gottesman, M. M., and S. V. Ambudkar. 2001. ABC Transporters and Human Disease. *J. Bioenerg. Biomembr.* 33:453–458.

[60] Grubmüller, H. 1996. Solvate 1.0. Theoretical Biophysics Group, Institut für Medizinische Optik, Ludwig-Maximilians-Universität München, München, Germany.

[61] Gullingsrud, J., and K. Schulten. 2003. Gating of mscl studied by steered molecular dynamics. *Biophys. J.* 85:2087–99.

[62] Gullingsrud, J., and K. Schulten. 2004. Lipid bilayer pressure profiles and mechanosensitive channel gating. *Biophys. J.* 86:3496–3509.

[63] Haliloglu, T., and I. Bahar. 1999. Structure-Based Analysis of Protein Dynamics: Comparison of theoretical Results for Hen Lysozyme With X-Ray Diffraction and NMR Relaxation Data. *Proteins.* 37:654–667.

[64] Hansen, F. Y., and G. H. Peters. 2003. Molecular Dynamics Simulations - from theory to Practice. Department of Chemistry, Department of Chemistry, The Technical University of Denmark, Lyngby, Denmark.

[65] Harasima, A. 1958. Molecular theory of surface tension. *Adv. Chem. Phys.* 1:203–237.

[66] Hermans, J., B. H. J. C., W. F. van Gunsteren, and J. P. M. Postma. 1984. A consistent empirical potential for water-protein interactions. *Biopolymers.* 23:1513–1518.

[67] Hess, B., H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije. 1997. LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.* 18:1463–72.

[68] Heyes, D. M. 1994. Pressure tensor of partial-charge and point-dipole lattices with bulk and surface geometries. *Phys. Rev. B.* 49:755–64.

[69] Higgins, C. F. 1992. ABC Transporters - From Microorganisms To Man. *Annu. Rev. Cell Biol.* 8:67–113.

[70] Hinsen, K., and G. R. Kneller. 1999. A simplified force field for describing vibrational protein dynamics over the whole frequency range. *J. Chem. Phys.* 111:10766–9.

[71] Holland, I. B., and M. A. Blight. 1999. ABC-ATPases, Adaptable Energy Generators Fuelling Transmembrane Movement of a Variety of Molecules in Organisms from Bacteria to Humans. *J. Mol. Biol.* 293:381–399.

[72] Horio, M., M. M. Gottesman, and I. Pastan. 1988. ATP-Dependent Transport of Vinblastine in Vesicles from Human Multidrug-Resistant Cells. *P. Natl. Acad. Sci. USA.* 85:3580–3584.

[73] Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD – Visual Molecular Dynamics. *J. Mol. Graphics.* 14:33–38.

[74] Irving, J. H., and J. G. Kirkwood. 1950. The statistical mechanical theory of transport processes. *J. Chem. Phys.* 18:817–29.

[75] Israelachvili, J. N. 1991. Intermolecular and surface forces. Academic Press.

[76] Israelachvili, J. N., S. Marcelja, and R. G. Horn. 1980. Physical principles of membrane organization. *Q. Rev. Biophys.* 13:121–200.

[77] Jähnig, F. 1996. What is the surface tension of a lipid bilayer membrane? *Biophys. J.* 71:1348–1349.

[78] James, S. R., R. A. Demel, and C. P. Downes. 1994. Interfacial hydrolysis of phosphatidylinositol 4-phosphate and phosphatidylinositol 4,5-bisphosphate by turkey erythrocyte phospholipase C. *Biochem. J.* 298:499–506.

[79] Jensen, M. Ø., and O. G. Mouritsen. 2004. Lipids do influence protein function-the hydrophobic matching hypothesis revisited. *Biochim. Biophys. Acta - Biomembranes.* 1666:205–226.

[80] Jensen, M. Ø., O. G. Mouritsen, and G. H. Peters. 2001. Dipolar and chain-linking effects on the rheology of grafted chains in a nanopore under shear at different grafting densities. *Phys. Rev. E.* 64:1–13.

[81] Jensen, M. Ø., O. G. Mouritsen, and G. H. Peters. 2004. The hydrophobic effect: Molecular dynamics simulations of water confined between extended hydrophobic and hydrophilic surfaces. *J. Chem. Phys.* 120:9729–9744.

[82] Jensen, M. Ø., O. G. Mouritzen, and G. H. Peters. 2004. Simulations of a membrane-anchored peptide: Structure, dynamics, and influence on bilayer properties. *Biophys. J.* 86:3556–3575.

[83] Jensen, T. R., M. Ø. Jensen, N. Reitzel, K. Balashev, G. H. Peters, K. Kjaer, and T. Bjørnholm. 2003. Water in contact with extended hydrophobic surfaces: Direct evidence of weak dewetting. *Phys. Rev. Lett.* 90:086101/1–086101/4.

[84] Jianbin He, Zhiyong Zhang, Yunyu Shi, and Haiyan Liu. 2003. Efficiently explore the energy landscape of proteins in molecular dynamics simulations by amplifying collective motions. *J. Chem. Phys.* 119:4005–4017.

[85] Jones, P. M., and A. M. George. 2002. Mechanism of ABC transporters: A molecular dynamics simulation of a well characterized nucleotide-binding subunit. *P. Natl. Acad. Sci. USA.* 99:12639–12644.

[86] Jones, P. M., and A. M. George. 2004. The ABC transporter structure and mechanism: perspectives on recent research. *Cell. Mol. Life Sci.* 61:682–699.

[87] Jorgensen, W. L., J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926–35.

[88] Jorgensen, W. L., and J. Tirado-Rives. 1988. The OPLS potential functions for proteins. energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.* 110:1657–1666.

[89] Kale, L., K. Schulten, R. Skeel, M. Bhandarkar, R. Brunner, A. Gursoy, N. Krawetz, J. Phillips, A. Shinozaki, and K. Varadarajan. 1999. NAMD2: Greater scalability for parallel molecular dynamics. *J. Comput. Phys.* 151:283–312.

[90] Kandt, C., W. Ash, and D. P. Tieleman. 2006. Setting up and running MD simulations of membrane proteins. *Methods.* (In Press).

[91] Karpowich, N. K., H. H. Huang, P. C. Smith, and J. F. Hunt. 2003. Crystal structures of the BtuF periplasmic-binding protein for vitamin

B12 suggest a functionally important reduction in protein mobility upon ligand binding. *J. Biol. Chem.* 278:8429–8434.

[92] Kechuan Tu, D. J. Tobias, J. K. Blasie, and M. L. Klein. 1996. Molecular dynamics investigation of the structure of a fully hydrated gel-phase dipalmitoylphosphatidylcholine bilayer. *Biophys. J.* 70:595–608.

[93] Kitao, A., and N. Go. 1999. Investigating protein dynamics in collective coordinate space. *Curr. Opin. Struc. Biol.* 9:164–169.

[94] Klauda, J. B., B. R. Brooks, A. D. MacKerell, Jr., R. M. Venable, and R. W. Pastor. 2005. An ab initio study on the torsional surface of alkanes and its effect on molecular simulations of alkanes and a DPPC bilayer. *J. Phys. Chem. B.* 109:5300–5311.

[95] Klauda, J. B., N. Kučerka, B. R. Brooks, R. W. Pastor, and J. F. Nagle. 2006. Simulation-based methods for interpreting x-ray data from lipid bilayers. *Biophys. J.* .

[96] Kučerka, N., Y. Liu, N. Chu, H. I. Petrache, S. Tristram-Nagle, and J. F. Nagle. 2005. Structure of fully hydrated fluid phase DMPC and DLPC lipid bilayers using X-ray scattering from oriented multilamellar arrays and from unilamellar vesicles. *Biophys. J.* 88:2626–2637.

[97] Lindahl, E., and O. Edholm. 2000. Mesoscopic undulations and thickness fluctuations in lipid bilayers from molecular dynamics simulations. *Biophys. J.* 79:426–33.

[98] Lindahl, E., and O. Edholm. 2000. Spatial and energetic-entropic decomposition of surface tension in lipid bilayers from molecular dynamics simulations. *J. Chem. Phys.* 113:3882–93.

[99] Lindahl, E., and O. Edholm. 2001. Molecular dynamics simulation of nmr relaxation rates and slow dynamics in lipid bilayers. *J. Chem. Phys.* 115:4938–4950.

[100] Lindahl, E., B. Hess, and D. van der Spoel. 2001. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J. Mol. Model.* 7:306–317.

[101] Locher, K. P. 2004. Structure and mechanism of ABC transporters. *Curr. Opin. Struc. Biol.* 14:426–431.

[102] Locher, K. P., and E. Borths. 2004. ABC transporter architecture and mechanism: implications from the crystal structures of BtuCD and BtuF. *FEBS Lett.* 564:264–268.

[103] Locher, K. P., A. T. Lee, and D. C. Rees. 2002. The E. coli BtuCD structure: A framework for ABC transporter architecture and mechanism. *Science.* 296:1091–8.

[104] Lu, G., J. M. Westbrooks, A. L. Davidson, and J. Chen. 2005. ATP hydrolysis is required to reset the ATP-binding cassette dimer into the resting-state conformation. *P. Natl. Acad. Sci. USA.* 102:17969–17974.

[105] Ma, J. 2005. Usefulness and Limitations of Normal Mode Analysis in Modeling Dynamics of Biomolecular Complexes. *Structure.* 13:373–380.

[106] MacCallum, J. L., and D. P. Tieleman. 2006. Computer simulation of the distribution of hexane in a lipid bilayer: Spatially resolved free energy, entropy, and enthalpy profiles. *J. Am. Chem. Soc.* 128:125–130.

[107] MacKerell, Jr., A. D., D. Bashford, M. Bellott, R. L. Dunbrack Jr, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher III, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus. 1998. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B.* 102:3586–3616.

[108] MacKerell, Jr., A. D., B. Brooks, C. L. Brooks III, L. Nilsson, B. Roux, Y. Won, and M. Karplus. 1998. CHARMM: The energy function and its parametrization with an overview of the program. *In* The Encyclopedia of Computational Chemistry. P. v. R. Schleyer, N. L. Allinger, T. Clark, J. Gasteiger, P. A. Kollman, H. F. Schaefer, III, and P. R. S. Schreiner, editors, volume 1. John Wiley & Sons, Chichester. 271–277.

[109] Maple, J. R. 1998. Force fields: A general discussion. *In* The Encyclopedia of Computational Chemistry. P. v. R. Schleyer, N. L. Allinger, T. Clark, J. Gasteiger, P. A. Kollman, H. F. Schaefer, III, and P. R. S. Schreiner, editors, volume 1. John Wiley & Sons, Chichester. 1015–1024.

[110] Marrink, S. J., and A. E. Mark. 2001. Effect of undulations on surface tension in simulated bilayers. *J. Phys. Chem. B.* 105:6122–27.

[111] Marsh, D. 1996. Lateral pressure in membranes. *Biochim. Biophys. Acta.* 1286:183–223.

[112] Marsh, D. 1997. Renormalization of the tension and area expansion modulus in fluid membranes. *Biophys. J.* 73:865–869.

[113] Miyamoto, S., and P. A. Kollman. 1992. SETTLE: an analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* 13:952–62.

[114] Monticelli, L., D. P. Tieleman, and G. Colombo. 2005. Mechanism of helix nucleation and propagation: Microscopic view from microsecond time scale MD simulations. *J. Phys. Chem. B.* 109:20064–20067.

[115] Mouritsen, O. G. 2005. Life - as a matter of fat. The emerging science of lipidomics. Springer-Verlag, Berlin Heidelberg.

[116] Nagle, J. F., and S. Tristram-Nagle. 2000. Structure of lipid bilayers. *Biochim. Biophys. Acta*. 1469:159–95.

[117] Nagle, J. F., R. Zhang, S. Tristram-Nagle, W. Sun, H. I. Petrache, and R. M. Suter. 1996. X-ray structure determination of fully hydrated L_α phase dipalmitoylphosphatidylcholine bilayers. *Biophys. J.* 70:1419–31.

[118] Neidhardt *et al.*, F. C. 1996. *Escherichia coli* and *Salmonella* - cellular and molecular biology, volume 1, 2nd Ed. ASM Press, Washington, D.C.

[119] Nosé, S., and M. L. Klein. 1983. Constant pressure molecular dynamics for molecular systems. *Mol. Phys.* 50:1055–76.

[120] Oloo, E. O., and D. P. Tieleman. 2004. Conformational transitions induced by the binding of MgATP to the vitamin B-12 ATP-binding cassette (ABC) transporter BtuCD. *J. Biol. Chem.* 279:45013–45019.

[121] Pastor, R. W., and S. E. Feller. 1996. Time scales of lipid dynamics and molecular dynamics. *In* Biological Membranes: A Molecular Perspective from Computation and Experiment. K. Merz, Jr., and B. Roux, editors. Birkhäuser, Boston. 3–30.

[122] Patra, M., M. Karttunen, M. T. Hyvonen, E. Falck, P. Lindqvist, and I. Vattulainen. 2003. Molecular dynamics simulations of lipid bilayers: major artifacts due to truncating electrostatic interactions. *Biophys. J.* 84:3636–45.

[123] Pearlman, D. A., D. A. Case, J. W. Caldwell, W. S. Ross, T. E. Cheatham, S. DeBolt, D. Ferguson, G. Seibel, and P. Kollman. 1995. AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comput. Phys. Commun.* 91:1–41.

[124] Rand, R. P., and V. A. Parsegian. 1989. Hydration forces betweeen phospholipid bilayers. *Biochim. Biophys. Acta*. 988:351–376.

[125] Reynolds, C. A., J. W. Essex, and W. G. Richards. 1992. Atomic charges for variable molecular conformations. *J. Am. Chem. Soc.* 114:9075–9079.

[126] Riordan, J. R., J. M. Rommens, B. Kerem, N. Alon, R. Rozmahel, Z. Grzelczak, J. Zielenski, S. Lok, N. Plavsic, J.-L. Chou, M. L. Drumm, M. C. Iannuzzi, F. S. Collins, and L.-C. Tsui. 1989. Identification of the Cystic Fibrosis Gene: Cloning and Characterization of Complementary DNA. *Science*. 245:1066–1073.

[127] Rowlinson, J. S., and B. Widom. 2002. Molecular Theory of Capillarity. Dover.

[128] Sachs, J. N., H. I. Petrache, and T. B. Woolf. 2003. Interpretation of small angle x-ray measurements guided by molecular dynamics simulations of lipid bilayers. *Chemistry and Physics of Lipids*. 126:211–223.

[129] Safran, A. 1994. Statistical Thermodynamics of Surfaces, Interfaces, and Membranes. Addison-Wesley Publishing Company.

[130] Schlenkrich, M., J. Brickmann, A. D. MacKerell, Jr., and M. Karplus. 1996. Empirical potential energy function for phospholipids: Criteria for parameter optimization and applications. *In* Biological Membranes: A Molecular Perspective from Computation and Experiment. K. Merz, Jr., and B. Roux, editors. Birkhäuser, Boston. 31–81.

[131] Schofield, P., and J. R. Henderson. 1982. Statistical mechanics of inhomogeneous fluids. *Proc. R. Soc. London, Ser. A*. 379:231–46.

[132] Seelig, A., and J. Seelig. 1974. The dynamic structure of fatty acyl chains in a phospholipid bilayer meassured by deuterium magnetic resonance. *Biochemistry*. 13:4839–4845.

[133] Singer, S., and G. Nicolson. 1972. The fluid mosaic model of the structure of cell membranes. *Science*. 175:720–731.

[134] Singh, U. C., and P. A. Kollman. 1984. An approach to computing electrostatic charges for molecules. *J. Comput. Chem.* 5:129–145.

[135] Skibinsky, A., R. M. Venable, and R. W. Pastor. 2005. Muscle and contractility - a molecular dynamics study of the response of lipid bilayers and monolayers to trehalose. *Biophysical J.* 89:4111–4121.

[136] Smart, O. S., J. M. Goodfellow, and B. A. Wallace. 1993. The Pore Dimensions of Gramicidin A. *Biophys. J.* 65:2455–2460.

[137] Sonne, J., F. Y. Hansen, and G. H. Peters. 2005. Methodological problems in pressure profile calculations for lipid bilayers. *J. Chem. Phys.* 122:124903.

[138] Stuart, S. J., Ruhong Zhou, and B. J. Berne. 1996. Molecular dynamics with multiple time scales: the selection of efficient reference system propagators. *Journal of Chemical Physics*. 105:1426–1436.

[139] Sukharev, S. I., P. Blount, B. Martinac, F. R. Blattner, and C. Kung. 1994. A large-conductance mechanosensitive channel in E. coli encoded by Msc[L alone. *Nature*. 368:265–268.

[140] Szleifer, I., D. Kramer, A. Ben-Shaul, W. M. Gelbart, and S. A. Safran. 1990. Molecular theory of curvature elasticity in surfactant films. *J. Chem. Phys.* 92:6800–17.

[141] Tajkhorshid, E., P. Nollert, M. Ø. Jensen, L. J. W. Miercke, J. O'Connell, R. M. Stroud, and K. Schulten. 2002. Control of the selectivity of the aquaporin water channel family by global orientational tuning. *Science*. 296:525–530.

[142] Tieleman, D., and H. Berendsen. 1998. A molecular dynamics study of the pores formed by Escherichia coli OmpF porin in a fully hydrated palmitoyloleoylphosphatidylcholine bilayer. *Biophys. J.* 74:2786–2801.

[143] Tieleman, D. P., and H. J. C. Berendsen. 1996. Molecular dynamics simulations of a fully hydrated dipalmitoylphosphatidylcholine bilayer with different macroscopic boundary conditions and parameters. *J. Chem. Phys.* 105:4871–80.

[144] Tieleman, D. P., and H. J. C. Berendsen. 1998. A molecular dynamics study of the pores formed by Escherichia coli OmpF porin in a fully hydrated palmitoyloleoylphosphatidylcholine bilayer. *Biophys. J.* 74:2786–801.

[145] Tieleman, D. P., J. L. MacCallum, W. L. Ash, C. Kandt, Z. Xu, and L. Monticelli. 2006. Membrane protein simulations with a united-atom lipid and all-atom protein model: lipid-protein interactions, side chain transfer free energies and model proteins. *J. Phys. - Condens. Mat.* .

[146] Tieleman, D. P., S. J. Marrink, and H. J. C. Berendsen. 1997. A computer perspective of membranes: molecular dynamics studies of lipid bilayer systems. *Biochim. Biophys. Acta.* 1331:235–70.

[147] Tieleman, D. P., I. H. Shrivastava, M. R. Ulmschneider, and M. S. P. Sansom. 2001. Proline-Induced Hinges in Transmembrane Helices: Possible Roles in Ion Channel Gating. *Proteins.* 44:63–72.

[148] Tirion, M. M. 1996. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys. Rev. Lett.* 77:1905–1908.

[149] van der Does, C., and R. Tampe. 2004. How do ABC transporters drive transport? *Biol. Chem.* 385:927–934.

[150] van der Ploeg, P., and H. J. C. Berendsen. 1982. Molecular dynamics simulation of a bilayer membrane. *J. Chem. Phys.* 76:3271–3276.

[151] van Gunsteren, W. F., and H. J. C. Berendesen. 1987. GROMOS87 Manual. Biomos, Groningen.

[152] Van Gunsteren, W. F., and A. E. Mark. 1998. Validation of molecular dynamics simulation. *J. Chem. Phys.* 108:6109–16.

[153] Varnik, F., J. Baschnagel, and K. Binder. 2000. Molecular dynamics results on the pressure tensor of polymer films. *J. Chem. Phys.* 113:4444–53.

[154] Venable, R. M., B. R. Brooks, and R. W. Pastor. 2000. Molecular dynamics simulations of gel ($L_{\beta'}$) phase lipid bilayers in constant pressure and constant surface area ensembles. *J. Chem. Phys.* 112:4822–4832.

[155] Vergani, P., S. W. Lockless, A. C. Nairn, and D. C. Gadsby. 2005. CFTR channel opening by ATP-driven tight dimerization of its nucleotide-binding domains. *Nature.* 433:876–880.

[156] Walton, J. P. R. B., and K. F. Gubbins. 1985. The pressure tensor in the inhomogeneous fluid of non-spherical molecules. *Mol. Phys.* 55:679–88.

[157] Walton, J. P. R. B., D. J. Tildesley, J. S. Rowlinson, and J. R. Henderson. 1983. The pressure tensor at the planar surface of a liquid. *Mol. Phys.* 48:1357–68.

[158] Wohlert, J., and O. Edholm. 2004. The range and shielding of dipole-dipole interactions in phospholipid bilayers. *Biophys. J.* 87:2433–2445.

[159] Woods, R. J., and R. Chappelle. 2000. Restrained electrostatic potential atomic partial charges for condensed-phase simulations of carbohydrates. *J. Mol. Struct. - Theochem.* 527:149–156.

[160] Zheng, W., and B. R. Brooks. 2005. Normal-modes-based prediction of protein conformational changes guided by distance constraints. *Biophys. J.* 88:3109–17.

7 Appendices

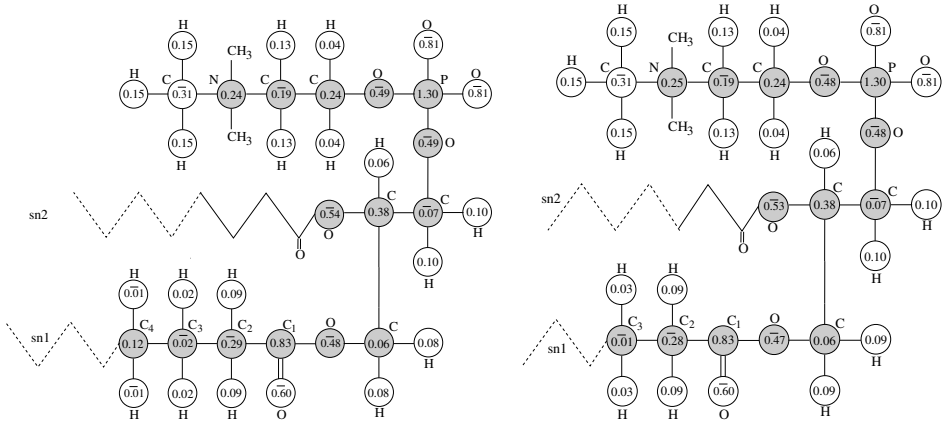
7.1 New partial Charges

Table 7.1 shows the average Mulliken and RESP charges obtained for dipentanoatephosphatidylcholine (DPePC). The Mulliken and RESP methods are denoted I and II, respectively. Also shown are the dipalmitoylphosphatidylcholine (DPPC) charges transferred from DPePC using transfer schemes A and B described in Sec. 3.2.2. The CHARMM27 partial charges are shown for reference. Figure 7.1 gives an overview of the new RESP charges listed

	Mulliken			RESP			CH27
	DPePC -	DPPC I.A	DPPC I.B	DPePC -	DPPC II.A	DPPC II.B	- III&IV
	partial charge (e)						
H1	0.245051	0.244909	0.245140	0.153367	0.150668	0.154454	0.25
C2	-0.369004	-0.369146	-0.368915	-0.309918	-0.312617	-0.308831	-0.35
N3	-0.568717	-0.568859	-0.568627	0.246550	0.243851	0.247637	-0.60
C4	-0.198324	-0.198466	-0.198235	-0.188332	-0.191031	-0.187245	-0.10
H5	0.244430	0.244288	0.244519	0.129583	0.126884	0.130670	0.25
C6	0.002913	0.002771	0.003002	0.241392	0.238693	0.242479	-0.08
H7	0.201919	0.201777	0.202008	0.041252	0.038553	0.042339	0.09
O8	-0.750061	-0.750203	-0.749972	-0.482624	-0.485323	-0.481537	-0.57
P9	1.616817	1.616675	1.616906	1.303270	1.300571	1.304356	1.50
O10	-0.831387	-0.831529	-0.831298	-0.810549	-0.813248	-0.809462	-0.78
C11	0.003425	0.003283	0.003514	-0.066394	-0.069093	-0.065307	-0.08
H12	0.187542	0.187400	0.187631	0.100650	0.097951	0.101737	0.09
C13	0.108801	0.108659	0.108891	0.378280	0.375581	0.379367	0.04
H14	0.232980	0.232838	0.233069	0.062906	0.060207	0.063992	0.09
C15	-0.025576	-0.025718	-0.025487	0.060502	0.057803	0.061589	-0.05
H16	0.214057	0.213915	0.214147	0.085715	0.083016	0.086801	0.09
O17	-0.656842	-0.656984	-0.656753	-0.472576	-0.475275	-0.471489	-0.34
C18	0.839348	0.839206	0.839437	0.832059	0.829360	0.833146	0.63
O19	-0.597498	-0.597640	-0.597408	-0.598219	-0.600918	-0.597132	-0.52
C20	-0.433166	-0.433308	-0.433076	-0.285440	-0.288139	-0.284353	-0.08
H21	0.203865	0.203723	0.203955	0.089737	0.087038	0.090824	0.09
C22	-0.314846	-0.314989	-0.314758	-0.012362	-0.015061	-0.011276	-0.18
H23	0.177125	0.176983	0.177215	0.026467	0.023769	0.027554	0.09
C24	-0.316755	-0.316897	-0.180000	0.118557	0.115859	-0.180000	-0.18
H25	0.161482	0.161340	0.090000	-0.007909	-0.010607	0.090000	0.09
C26	-0.486578	-0.180000	-0.180000	-0.267435	-0.180000	-0.180000	-0.18
H27	0.160867	0.090000	0.090000	0.063956	0.090000	0.090000	0.09
O28	-0.663024	-0.663166	-0.662935	-0.535437	-0.538136	-0.534350	-0.34

Table 7.1: Partial charges for DPePC and DPPC in different parameter sets. See Table 3.1 for more details.

in Table 7.1 after the transfer to DPPC. To increase clarity in the figure, the charges are shown with two decimals only. Due to round-offs, the molecules have a non-zero net charge, which is not the case when all six decimals are used (see Table 7.1).



(a) (b)

Figure 7.1: Overview of the new partial charges for the DPPC lipids as obtained from the RESP method. (a) Transfer scheme A where the atomic charges of C_1 to C_4 in the acyl chain were modified compared to CHARMM27. (b) Transfer scheme B where the atomic charges of C_1 to C_3 in the acyl chains were modified compared to CHARMM27. In both panels the atoms in the lipid 'backbone' are drawn in grey. The atoms in the three methyl groups on nitrogen have identical charges and for clarity only one methyl group is shown explicitly. The solid zigzag line in the sn2 chain indicates that these atoms have been assigned new charges, which are equal to the corresponding charges in the sn1 chain. For clarity, the atoms are only shown explicitely in the sn1 chain. The dashed lines in the sn1 and sn2 chains symbolize the rest of the acyl chain where the CHARMM27 charges were used. A bar indicates that the charge is negative. Note that due to the rounding of the charges in the figure leave DPPC with a non-zero net charge, which is not the case when all 6 decimals are included (see Table 7.1).

7.2 Reciprocal space pressure profile contribution

For a rectangular box, the xx element of the configurational pressure contribution can be written as [25, 119]:

$$P^{xx}V = -\frac{\partial U}{\partial L^x}L^x, \tag{7.1}$$

where L^x is the box length in the x direction and V is the simulation box volume. U is the potential energy. The energy of the ith particle from the reciprocal space part of the Ewald sum $U_{K,i}$ can be expressed as [68]:

$$U_{K,i} = \frac{q_i}{V\epsilon_0}\sum_{\mathbf{k}_n\neq\mathbf{0}} Q(k_n)\mathrm{Re}\left[\exp(-\mathrm{i}\mathbf{k}_n\cdot\mathbf{r}_i)S(\mathbf{k}_n)\right], \tag{7.2}$$

satisfying $U_K = \frac{1}{2}\sum_i U_{K,i}$ where U_K is the first term in Eq. (4.7). The reciprocal pressure contribution from this particle therefore reads

$$\begin{aligned} P^{xx}_{K,i}V = &-\frac{\partial U_{K,i}}{\partial L^x}L^x = \\ &-\frac{q_i}{\epsilon_0}\frac{\partial}{\partial L^x}\Big\{\frac{1}{V}\sum_{\mathbf{k}_n\neq\mathbf{0}} Q(k_n)\mathrm{Re}\left[\exp(-\mathrm{i}\mathbf{k}_n\cdot\mathbf{r}_i)S(\mathbf{k}_n)\right]\Big\}L^x, \end{aligned} \tag{7.3}$$

and therefore

$$\begin{aligned} \frac{-P^{xx}_{K,i}V\epsilon_0}{q_iL^x} = &\frac{\partial(1/V)}{\partial L^x}\sum_{\mathbf{k}_n\neq\mathbf{0}} Q(k_n)\mathrm{Re}\left[\exp(-\mathrm{i}\mathbf{k}_n\cdot\mathbf{r}_i)S(\mathbf{k}_n)\right] \\ &+\frac{1}{V}\sum_{\mathbf{k}_n\neq\mathbf{0}}\frac{\partial Q(k_n)}{\partial L^x}\mathrm{Re}\left[\exp(-\mathrm{i}\mathbf{k}_n\cdot\mathbf{r}_i)S(\mathbf{k}_n)\right] \\ &+\frac{1}{V}\sum_{\mathbf{k}_n\neq\mathbf{0}} Q(k_n)\frac{\partial\mathrm{Re}\left[\exp(-\mathrm{i}\mathbf{k}_n\cdot\mathbf{r}_i)S(\mathbf{k}_n)\right]}{\partial L^x}. \end{aligned} \tag{7.4}$$

Since $V = L^xL^yL^z$, the first term differentiates to:

$$\frac{\partial(1/V)}{\partial L^x} = -\frac{1}{VL^x}. \tag{7.5}$$

The second term differentiates to:

$$\begin{aligned} \frac{\partial Q(k_n)}{\partial L^x} &= \frac{\partial}{\partial L^x}\Big(\frac{\exp(-k_n^2/4\kappa^2)}{k_n^2}\Big) \\ &= \frac{\partial}{\partial k_n^x}\Big(\frac{\exp(-k_n^2/4\kappa^2)}{k_n^2}\Big)\frac{\partial k_n^x}{\partial L^x} \\ &= \Big(\frac{2k_n^x}{4\kappa^2}\frac{\exp(-k_n^2/4\kappa^2)}{k_n^2} + 2k_n^x\frac{\exp(-k_n^2/4\kappa^2)}{k_n^4}\Big)\frac{2\pi n^x}{(L^x)^2} \\ &= \frac{2k_n^xk_n^x}{L^x}Q(k_n)\left[\frac{1}{4\kappa^2}+\frac{1}{k_n^2}\right]. \end{aligned} \tag{7.6}$$

Since $k_n^x r_i^x = \frac{2\pi n^x}{L^x} r_i^x = 2\pi n^x s_i^x$, where s_i^x is the scaled x coordinate of particle i, the third term is zero as scaled coordinates are independent of of the box dimensions. Inserting Eqs. (7.5) and (7.6) into Eq. (7.4) we recover Eq. (4.12) of Sec. 4.2.3.

7.3 prespro1.1 user guide

This is a brief user guide for the prespro1.1 code.

About the code

The code calculates the pressure profile in the z-direction according to three different methods: (1) The Irving-Kirkwood method using a cutoff for the electrostatic interactions. (2) The Harasima method using a cutoff for the electrostatic interactions. (3) Harasima-Ewald method which includes the contribution to the pressure profile coming from the reciprocal space part of the Ewald summation (see Chapter 4 or Ref. [137]).

Installation

The distribution contains the files

- PPmain.C/h (main)
- Bonded.C/h (calculates bonded contributions)
- Nonbonded.C/h (calculates nonbonded contributions)
- profile.h (some constants)
- Complx.h (a complex class)
- Tensor.h (a tensor class)
- Makefile
- dcdreader.C/h (some dcd-reading functions)
- readdcd.C/h (some dcd-reading functions)

The files dcdreader.C/h readdcd.C/h are taken from CatDCD 2.0 without modifications. CatDCD 2.0 is no longer available from http://www.ks.uiuc.edu/Development/MDTools/ and therefore these files are included in this distribution. The CatDCD 2.0 files comes with the following copyright notice:

```
************************************************************************
* University of Illinois Open Source License                           *
* Copyright 2003-2005 Theoretical and Computational Biophysics Group,  *
* All rights reserved.                                                 *
*                                                                      *
* Developed by:          Theoretical and Computational Biophysics Group *
*                        University of Illinois at Urbana-Champaign    *
*                        http://www.ks.uiuc.edu/                       *
************************************************************************
```

The above listed files only make up a part of the pressure profile code. Due to license agreements we cannot distribute the rest of the code, but it is available from `http://www.ks.uiuc.edu/Development/MDTools/`. From this site download Mindy 1.0, uncompress, and copy all files except

```
Integrator.C ComputeBonded.C/h ComputeNonbonded.C/h Makefile
```

to the `./src/` directory. Go to the `./src/` directory. The default force field expected by Mindy is X-plore. This is easily changed to CHARMM by adding the line

```
#include "profile.h"
```

in line 7 of `Parameters.C` and replacing the lines

```
paramType = paraXplor;
read_parameter_file(psf);
#ifdef MIN_CHARMMFORMAT
  paramType = paraCharmm;
  read_charmm_parameter_file(psf);
#endif
```

(i.e. lines 148-153) by the following:

```
#ifndef MIN_CHARMMFORMAT
  paramType = paraXplor;
  read_parameter_file(psf);
#endif

#ifdef MIN_CHARMMFORMAT
  cout << "Using CHARMM parameters." << endl;
  paramType = paraCharmm;
  read_charmm_parameter_file(psf);
#endif
```

Note that the code has not been tested using X-plore parameters. In the `./src/` directory type "make" and you will get an binary file called `prespro_g++`.

Usage

Once compiled, you will need six input files before you can run the pressure profile analysis:

1. pdb file
2. psf file

3. dcd trajectory file
4. parameter file
5. temperature file; one column of temperatures. One temperature per dcd-frame.
6. box file; three columns with the x, y and z dimensions of the box. One line per dcd-frame

It is crucial that the frames in the box-file matches the frames in the dcd-file exactly in order for the periodic boundary conditions to be correct. The pressure profile program is executed as

```
prespro_g++     ncm pdb psf par dcd first last
                skip nslabs kappa kmaxsq cutoff
                swd box temp odir
```

The command line input parameters are:

- `ncm`: the number of atoms in the determination of the center for the pressure profile. The program only considers the first `ncm` atoms, which should therefore be the lipids.
- `pdb`: the pdb file
- `psf`: the psf file
- `par`: the parameter file
- `dcd`: the dcd file
- `first`: the first frame to read
- `last`: the last frame to read
- `skip`: stride (skip)
- `nslabs`: the number of slabs in the pressure profile calculation
- `kappa`: the shift parameter for Ewald summation. Increasing `kappa` shifts the Ewald sum towards reciprocal space and therefore when increasing `kappa`, `kmaxsq` should be increased and `cutoff` can be reduced.
- `kmaxsq`: squared cutoff in reciprocal space (i.e. max of $\texttt{kx}^2+\texttt{ky}^2+\texttt{kz}^2$). The value of `kmaxsq` depends on `kappa` and the box dimension.
- `cutoff`: the cutoff in real space

- `swd`: switch distance in real space
- `box`: the box file
- `temp`: the temperature file
- `odir`: the output directory

Output

The output file names are put together from a prefix (`p`), the interaction type (`i`) and the calculation method (`m`), as: `p_i_m.dat`.

Prefix (`p`)

There are five classes of pressure output indicated by the prefix `p`:

1. `tpxx` = xx element of the average simulation pressure
2. `tlpt` = time resolved local pressure tensor
3. `appt` = average pressure tensor profile
4. `appN` = average normal pressure profile (column 9) of 3.
5. `appL` = average lateral pressure profile (xx+yy)/2 (column 1 and 5) of 3.

`tpxx`: The xx element of the total pressure (in bar) for each type of interaction (`i`). These output files are mostly useful for debugging and for adjusting the Ewald parameters. Format:

```
ptot_xx(frame_1)
ptot_xx(frame_2)
:
:
ptot_xx(frame_n)
```

`tlpt`: These output files are the complete, but confusing output. May be useful for statistical analyses. However, we found that averaging the pressure profile over 250 ps gives uncorrelated profiles as well a decent number of profiles say from a 20 ns trajectory. Thus, the average pressure profile from a 250 ps trajectory piece might be more convenient (see below). Format: Each line holds the nine pressure tensor elements in the slabs `1,2...nslabs` (i.e. 9 × `nslabs` columns). Each line is a frame(1,2...n)

```
xx(1) yx(1) zx(1) xy(1) yy(1) xz(1) yz(1) zz(1) xx(2) yx(2) zx(2) .... zz(nslabs)
xx(1) yx(1) zx(1) xy(1) yy(1) xz(1) yz(1) zz(1) xx(2) yx(2) zx(2) .... zz(nslabs)
:
:
xx(1) yx(1) zx(1) xy(1) yy(1) xz(1) yz(1) zz(1) xx(2) yx(2) zx(2) .... zz(nslabs)
```

appt: One line holds the nine element of the average pressure profile for the trajectory in one slab. Thus, there are nslabs lines. Format:

```
pxx(1) pyx(1) pzx(1) pxy(1) pyy(1) pxz(1) pyz(1) pzz(1)
pxx(2) pyx(2) pzx(2) pxy(2) ....
:
:
pxx(nslabs) pyx(nslabs) pzx(nslabs) pxy(nslabs)
```

appN: One line contains the average normal pressure in one slab. Thus, there are nslabs lines that should be identical to column 9 of the corresponding appt-file. Format:

```
pzz(1)
pzz(2)
:
:
pzz(nslabs)
```

appL: One line contains the average lateral pressure in one slab. Thus, there are nslabs lines that should be identical to the average of columns 1 and 5 of the corresponding appt-file. Format:

```
pL(1)
pL(2)
:
:
pL(nslabs)
```

Interaction types (i)

There are 11 interaction types (i):

1) bnd = valence bond
2) van = valence angle
3) dih = dihedral
4) imp = improper
5) vdw = van der Waals
6) ec = Electrostatic using cutoff (no shifting)
7) ecs = Electrostatic using cutoff and shift function
8) esr = real space electrostatics (Ewald)

9) `esk` = reciprocal space electrostatics (Ewald)

10) `kin` = kinetic pressure

11) `tot` = 1) + 2) + 3) + 4) + 5) + 8) + 9) + 10)

Methods (`m`)

There are three methods for calculating the pressure profile (`m`): 1) ik = Irving-Kirkwood 2) ha = Harasima 3) The HE-method is identical to the standard Harasima method for all interactions except for the electrostatics. Thus, for the `ec` and `ecs` interactions the `ha` indicates the use of a cutoff whereas when dealing with `esr` and `esk` `ha` indicates the HE-method.

In addition to the pressure output, the slab width in each frame is written to `slab_width.dat`. This is useful when converting slab numbers to actual positions.

In the `./test/` directory the `test.sh` script will run a simple test-run and compare the output to the original simulation output. The test script assumes that you installed 'catdcd.LINUX' and that it's location is in your path. The pdb and psf files are only meant for testing - do not use them for scientific purposes!

Tricks, limitations and issues

- The total length of any `outdir + outfile` string should be less than 50 characters

- You'll find a significant speedup if you're using the Intel C++ compiler instead of gcc

- The code does not run in parallel but you can use CatDCD to chop up your trajectory in suitable bites and run each piece on a single processor. This is also useful for statistical analyses. I used 250 ps per piece, which gave uncorrelated pressure profiles

- The total force array returned to `PPmain.C` should not be trusted

- Don't trust the off diagonal elements of the Harasima tensor

- The file `appN_esk_ha.dat` contains zeros since the reciprocal space contribution to the normal pressure cannot be calculated in the HE-method

- The code has not been tested using X-plore parameters

- You should use an update frequency for the nonbonded interactions of 1 in NAMD. Setting the update frequency to e.g. 2 will give large artifacts in the pressure profile: Especially the aqueous regions seems

to be very sensitive and I find a lateral (and normal) pressure of plus several thousand bar in these regions when using a update frequency of 2! This artifact seems to be isotropic (at least in the aqueous regions) and therefore it almost cancels in `pN-pL`. Reducing the update frequency to 1 completely eliminates this artifact and the local pressure in bulk water is close to 1 bar. Thus `Nonbondedfreq=1`. It seems OK to use `FullElectFreq=4`.

- I didn't implement twin range cutoff so if you want to calculate the van der Waals and electrostatic interactions with different cutoff you have to run the whole thing twice and combine the relevant output files. Sorry about that!

- Always check if the `prespro_g++` energies are the same as the energies from the simulation. You will probably never get an exact match but if there are big deviations some common user errors are:

  – The cutoff and the switch distance are not the same as in the simulation

  – You may need to adjust some of the Ewald parameters: `kappa`, `kmaxsq`

  – Ensure that the box-file and the dcd file are "aligned"

Versions

`v1.1`
- No bugs found since `v1.0` , but polished the code a bit
- Made output file names more systematic as to allow other people than the authors to understand the output
- Added the `appN` and `appL` files to output
- Removed calculation and output of P^2_{ab}
- The zz-component of Harasima tensor is now equal to the zz-element of the IK-tensor
- Moved most constants to `profile.h`
- Output the slab width to `slabwidth.dat`
- Now we're also reading the first frame in the frame loop.

`v1.0` The version used in Ref. [137]